Psychometrika

CONTENTS

FACTOR ANALYSIS AND PHYSICAL TYPES - - - - 171
CYRIL BURT

MULTIPLE BI-SERIAL AND MULTIPLE POINT BI-SERIAL CORRELATION - - - - - - - - - - - - 189
ROBERT J. WHERRY

CONFIRMATION AND CLARIFICATION OF PRIMARY PERSONALITY FACTORS - - - - - - - - 197
RAYMOND B. CATTELL

THE CONSISTENCY OF THE ITEMS OF AN ACTIVITY PREFERENCE BLANK - - - - - - - - 221
H. M. FOWLER

REVIEWS - - - - - - - - - - - - - - - - - 233

VOLUME TWELVE SEPTEMBER 1947 NUMBER THREE

PSYCHOMETRIKA—VOL. 12, NO. 3
SEPTEMBER, 1947

FACTOR ANALYSIS AND PHYSICAL TYPES

CYRIL BURT
UNIVERSITY COLLEGE, LONDON

Physical measurements, obtained from several different age-groups and national groups, have been correlated and factorized. The largest of the investigations was based on 2,400 British airmen in the R.A.F., and the most recent included as many as 17 traits. The methods of factorization employed were those worked out in early investigations of mental and scholastic tests, and differed in minor respects from those more recently used by Thurstone in his re-analysis of some of our data. The factor patterns reveal a remarkable constancy from age to age and from one national group to another. The general factor accounts for about the same proportion of the variance as the general factor in cognitive tests. The largest of the bipolar factors classifies traits into longitudinal and transverse, thus in some measure confirming the traditional classification of body-types.

Introduction

1. *Problem.* The study of the factors underlying somatic or physical measurements appears to have reached much the same position as that which obtained in the study of mental or psychological factors about thirty years ago. In the early days of factor analysis, when the method was applied almost exclusively to cognitive tests, three main rival hypotheses were put forward: (i) that the facts could be accounted for by a single, dominant "general factor," the influence of more specialized factors being "vanishingly minute" (Spearman, 1905); (ii) that *both* general *and* group factors were required (Burt, 1911); (iii) that a general factor was superfluous and that overlapping group factors would suffice (Thomson, 1919). Spearman's theory was epitomized as essentially a "theory of the absence of group factors"; Thomson, on the other hand, was equally definite that "there is no such thing as general ability." In later writings, however, Spearman grew less and less averse to group factors and Thomson more tolerant of a general factor, provided such factors were defined in consonance with their own respective views.

A similar controversy seems now to have developed in regard to the time-honored problem of physical types and the temperamental tendencies supposed to accompany them; and the main object of this paper is to urge that here, too, the most appropriate hypothesis is

one which impartially admits both group factors and a general factor, and which consequently seeks, not to rule out one or the other by a ruthless application of Occam's razor, but rather to assess the relative influence of each. The issue is raised most clearly by Prof. Thurstone's recent article (15), in which he has re-factorized a table of correlations between physical measurements calculated and factorized by one of my research-students (10), and expresses the hope that his discussion may be "provocative of more intensive studies of larger sets of anthropometric measurements." In reply I want first of all to show that a number of such studies have already been made, and seem, so far, to support the view I have suggested; secondly, to illustrate in greater detail the methods adopted, in the hope of meeting some of Prof. Thurstone's objections; and finally to examine, and if possible to reconcile, the differences between the results obtained by the two different procedures.

2. *The Psychological Interest of Physical Factors.* An introductory word or two is needed to explain why the analysis of physical measurements is thought worthy of the psychologist's attention. (i) In Great Britain, at the time when I first started work on individual psychology over forty years ago, the notion of physical types, each with its distinctive temperamental traits, was very much to the fore: it was strongly supported by the survivors of the faculty school; it was taught by many leading educationists, anthropologists, and medical authorities; and it was still more actively popularized by the phrenologists, who, particularly in North Britain, still possessed considerable academic influence. The scheme of types then most widely accepted was similar to that now largely associated with Kretschmer, and was derived from much the same sources, namely, the traditional version of the Galenic theory, as modernized by Addison, Hutchinson, Laycock, and Stewart in this country, and by Carus, di Giovanni, Sigaud, Beneke, and Bauer abroad. A tentative application of factor analysis to emotional assessments appeared partly to confirm the classification of temperaments (in the form of tendencies rather than types); and it was thought worth while to attempt a concurrent investigation on the corresponding classification of physical types [(5) and refs.]. (ii) Unless one is to accept the Cartesian doctrine that body and mind are wholly disparate, it is natural to expect that the same type of factor pattern will emerge from correlations between physical traits as from those between mental traits. This is in accordance with Prof. Thurstone's view, since he finds "simple structure in both." Spearman, on the other hand, who inclined to a dualistic view of the body-mind problem, held that the factorial constitution of the human body was wholly distinct from that of the mind, and that (except in cases

of organic mental disease) little light could be thrown by the one problem upon the study of the other. (iii) But in any case, since physical measurements (and to some extent their causes) are more accurately known than mental, it has always seemed to me that the study of factorial techniques might best begin by factorizing the more concrete and more readily visualizable traits of the body. This approach, I believe, is as helpful for the advanced theorist as for the elementary student; and my own students have in consequence always been advised to start their acquaintance with factorial methods by correlating and analyzing their own physical measurements.* (iv) Quite apart from its possible interest for psychologists, the factorial study of physical characteristics has other uses, both theoretical and practical, though these scarcely concern us here. From a theoretical standpoint the technique should prove instructive to anthropologists and biologists, who have frequently to deal with the classification of complex overlapping physical types. From a practical standpoint it should help to simplify certain problems of bodily measurement with which manufacturers and others are confronted both in war and in peace (e.g., the best sizes and shapes in which to issue uniforms and what should be the number of each, the way in which variation in size and shape of the body may affect the manufacture of tanks, aeroplanes, etc., into which the human body has to fit, etc.); indeed, it has been the emergence of such problems during the recent war that has supplied our most extensive collection of data.

Methods

I will begin by briefly explaining how our procedure differs from that of Thurstone, illustrating the chief differences by reference to the example he has reworked. In analyzing any given table of correlations our practice is to start with what I have termed the method of "simple summation." The formula which I gave for this purpose in 1917 (4, 53) is identical with Thurstone's "centroid formula" (13, 94); and, although our devices for determining the communalities and changing the signs of negative residuals are not quite the same as his, the changes rarely introduce any wide discrepancies into the resulting factor-saturations. The saturations (or "loadings") thus obtained we regard as merely a convenient approximation to those which would be obtained by the method of "weighted summation," when applied to the correlation matrix with the same communalities:

* In London all Honours students are required to take a "subsidiary subject" for their degree examinations; and for those whose "main subject" is psychology, anthropology is the commonest "subsidiary." Hence a large number are familiar with the techniques of physical anthropology.

the latter yields the series of best fitting factors, in accordance with the principle of least squares (6, 320f.). But in practice we seldom find it necessary to determine these weighted factors more precisely, except where the question at issue itself requires a rigorous assessment of the exact degree of fit.

1. *General Factors.* The first factor obtained by this procedure represents the factor that would best account for the observed data and correlations, were we limited to a single factor only. It is the component that furnishes the largest contribution to the total variance, and is thus the factor of greatest value for purposes of practical prediction. Spearman's interest centered mainly in the presence or absence of this general factor: he believed he could demonstrate its presence in all mental measurements and its absence from bodily measurements; the fact that I had found a general factor in physical measurements as well he attributed to the use of a faulty formula and the neglect of the tetrad difference criterion. If eventually confirmed for physical measurements, this general factor will presumably be interpreted as a factor of size or "isometric growth." Our first problem is therefore to determine its presence and measure its amount.

2. *Bipolar Analysis.* Our next problem is to examine the evidence for an antithetical pair of body types. To do this we have to eliminate the size-factor. When this is removed, the summation method yields what I call a "bipolar" factor, dividing the traits into two contrasted groups, distinguished by opposite signs: as we shall see in a moment, one group consists almost invariably of longitudinal traits and the other of transverse or circumferential traits.

3. *Group Factor Analysis.* Although bipolar factors provide the most direct procedure for confirming alleged dichotomous classifications, nevertheless, to study the relations between the tendencies disclosed, it is convenient to re-analyze the observed correlations into group-factors of greater or lesser generality. For this purpose we rely on a modified version of the same formula as before. The principle underlying the working procedure is to take either the bipolar factors or the intercolumnar product-moments as a preliminary clue to the groupings and then to sum the observed correlations within the several groups: the summation formula is applied to the reduced matrix consisting of these summed correlations, with the diagonal values estimated by successive approximation as before. (6, 477 and refs.).

Thurstone achieves much the same end by a more indirect approach, namely, by rotating his centroid factors until he has secured a "simple structure." But the requirements of simple structure are

specified a priori. Our method, on the other hand, attempts to deduce the group factor pattern afresh for each correlation matrix by using the evidence supplied by the initial bipolar analysis. This, we believe, avoids the conscious or unconscious influence of preconceived theories, which might affect the choice of axes during uncontrolled rotations. Thus, with Hammond's data (15, 19, Table 2) the plus and minus signs of the first bipolar factor (Thurstone's Factor II) group the traits into longitudinal and non-longitudinal. The second bipolar factor (Factor III) then subdivides the non-longitudinal into a Breadth group and a Head group, and the longitudinal group into Height measurements and Hand measurements; but for this latter subdivision (Height and Hand) the centroid saturations are so low that we must regard it as at most supplementary.

We begin, therefore, by assuming three main group factors only (Longitudinal, Transverse, and Head). With Hammond's table this at once yields a tolerably good fit to the observed coefficients. Two significant residuals remain, viz., for Sitting Height and Standing Height and for Hand breadth and Hand length, respectively. Accordingly we may tentatively subjoin a pair of factor saturations for each of these.

With data sufficiently extensive to warrant the additional labor, the complete procedure would be to use successive approximation, and to continue both the "bipolar" and the "group" analyses until the residuals are equally negligible. The correlation table reconstructed from the group factors is then factorized by the summational (centroid) method, and an exactly orthogonal transformation matrix can thus be determined: to shorten the working the factor saturations may first be summed by groups. The final set of overlapping group factors is obtained by applying this transformation matrix to the *original* matrix of summational factors.

Graphical methods, we believe, are not very trustworthy for executing rotations, except in rough preliminary explorations. But the various methods of graphical projection are extremely useful for exhibiting the relations of factors diagrammatically: the stereographic (with a specially printed net) is the most helpful for general purposes (including practical calculations); the globular easiest for the beginner; the gnomonic for making tentative rotations.

The saturations eventually obtained for Hammond's data by this procedure are shown below in Table 1. For comparison I subjoin the "centroid" factor-saturations given by Thurstone: the figures differ only slightly from those reached by Mr. Hammond with the summational method. To avoid confusing the pattern, saturations devoid of

TABLE 1
Factor Saturations for 12 Physical Traits: 100 Male Adults

Traits	A. Centroid Factors				B. Group Factors					
	I	II	III	IV	Size	Lept.	Pach.	Head	(Hand)	(Height)
1. Standing Height	.72	—.55	—.15	—.11	.506	.785	—	—	(.179)	(.337)
2. Sitting Height	.60	—.34	—.25	—.48	.524	.292	—	—	—	(.667)
5. Span	.79	—.35	.02	.03	.755	.488	—	—	(.202)	—
11. Hand Length	.78	—.41	.01	.23	.703	.452	—	—	(.326)	—
12. Hand Breadth	.58	—.15	.38	.31	.346	(.113)	.365	—	(.867)	—
3. Shoulder Breadth	.51	.17	.22	—.06	.466	—	.354	—	—	—
6. Chest Breadth	.60	.32	.24	—.16	.491	—	.547	—	—	—
7. Chest Depth	.34	.34	.29	—.26	.157	—	.641	—	—	—
4. Hip Breadth	.59	.21	.36	—.11	.470	—	.529	—	—	—
9. Head Breadth	.28	.29	—.36	.32	.213	—	—	.546	—	(—.121)
8. Head Length	.45	.31	—.30	.04	.423	—	(.106)	.452	—	—
10. Head Height	.23	.16	—.49	.19	.108	(.101)	—	.606	—	—
Thurstone's Factor	...	...	...	...	...	B	C	A	D	...

significance have been omitted from our list, and those of low significance bracketed.* This table was based on 100 persons only. But earlier and later studies made in our department—the latter based on far larger samples—show that the methods described are applicable to data that are susceptible of much more precise analysis, and that the group factor pattern here obtained is observable in numerous other investigations (7, 84-5, 1, 32-156).

4. *Criteria for Constancy of Factors.* In order to determine how far the factors reached with different samples of persons or with different selections of traits are really the same, some formal criterion seems essential. If we define a factor as the same when its direction-cosines are the same (regardless of the size of the factor variance), then its saturations must retain, not the same absolute values, but the same proportionate values. This, as I have shown elsewhere, leads at once to convenient criteria (6, 41 and refs.). The most rigorous test assumes that the factors for identification are those reached by weighted summation, and leads to what I termed

* Although the evidence for the supplementary factor-saturations is rather slender, they conform with what we might anticipate on other grounds, and the overlapping agrees with what has commonly been observed with this particular population (Irish). Even the small negative correlation between the height factor and head breadth is in keeping with the known fact that brachycephalic persons are commonly short. As noted in the text, Thurstone's rotated factor-pattern also includes a Hand-factor, but not a Height-factor, for which the residual would seem to be far more significant: since Standing Height necessarily includes Sitting Height and none of the other traits include it completely, there must be a group factor entering into these two.

the "symmetry criterion." With group factors a modified procedure is necessary. The simplest is to calculate $(G_1'G_1)^{-1}\ G_1'G_2$, where G_1 and G_2 denote the group factor matrices obtained from two successive researches, G_1 indicating by convention the matrix with the lower values. Now, if the factors are identical in the sense above defined, $G_2 = G_1D$, where D is a diagonal matrix. Hence the foregoing product should yield a diagonal matrix.

By way of illustration, let us take the five traits in Hammond's table which reappear in nearly the same form in the larger battery of measurements as obtained from 528 men and tabulated below (Table 4), viz., Standing Height, Sitting Height, Shoulder Breadth, Breadth at Hip or Waist and Chest Girth (obtained by combining Chest Breadth and Chest Depth). The above product-formula then yields the following figures.

TABLE 2
Diagonal Criterion

Factors	Size	Lept.	Pach.
Size	.982	—.134	—.046
Leptosomic	.244	.923	.034
Pachysomic	.273	.106	1.014

To measure the degree of diagonality we can take the expression $1 - \frac{S_o}{(r-1)S_d}$, where r is the number of factors, S_d the sum of the diagonal elements (the "trace"), and S_o the sum of all the other elements, taken regardless of sign. With the above figures this formula gives 0.857; so that the agreement is reasonably good.

If we want to test the constancy of each factor independently, then the proper criterion of proportionality consists in what I term the 'unadjusted correlation' between the two sets of factor-saturations.* When, as with bipolar factors obtained by simple summation, the sum of each column of saturations is zero, this reduces to the ordinary coefficient of correlation. Accordingly, there will be no great loss of accuracy and considerable gain in intelligibility, if we substitute for all purposes the ordinary coefficient calculated in the usual way. This accordingly will be employed in what follows.

* This is the ratio of the product-sum of the figures taken as they stand to the geometrical mean of their respective square-sums, without adjusting for the fact that the figures as they stand may not be deviations about a mean of zero (cf. 6, 342). This procedure has the advantage of yielding a criterion which is the same whichever set of factors we take as G_1. To test the significance of factor-correlations the chi-squared test may be employed. If we seek to take into account the effect of a multivariate selection of the population, we are led to more elaborate expressions, a little too cumbersome for practical use.

5. *Comparison with Thurstone's Analysis.* Thurstone's analysis "seems to indicate four primary factors . . . concerned with the development of (a) head size, (b) bone length, (c) transverse measures or girth, (d) a factor which may be concerned with size of extremities." So far our findings are in close agreement. The chief difference consists in the absence from Thurstone's pattern of any general factor (cf. 15, Table 4). This omission has only become possible because his oblique factors themselves have positive correlations; indeed, the correlation between his length and girth factors amounts to 0.38.

Now I venture to think that a procedure which simply substitutes correlated factors for correlated traits fails to bring out one of the most important points deducible from all such tables, namely, that human bodies grow at much the same relative rate in every direction, and therefore, when growth is over, exhibit much the same general shape. This fact is of obvious theoretical interest: it confirms for human beings what has been noted for so many other species—namely, that isometric growth dominates over allometric. On taking repeated measurements for the same children, it is easy to show that in most cases the individual's "size quotient," as it might be called (i.e., his general factor measurement when expressed as a ratio to the average measurement for all children of the same age), remains remarkably constant throughout the years of growth. It is thus in many ways analogous to the intelligence quotient: it has at least as high a predictive value: like the I.Q., it yields a correlation of about .5 for sibs;* and like the I.Q., it is lower among children from poorer homes and from larger families. But the factor is equally important for practical purposes. Thus, in making certain physiological estimations (e.g., blood volume, basal metabolism) weight is at present commonly used as an indication of body-size, but it would seem far more accurate to estimate it by a general factor measurement calculated from a regression equation. Or again consider one of the war-time problems already mentioned — that of the service-tailor manufacturing ready-made uniforms: the first thing he wants to know is how far it would be safe to make most uniforms of the same shape, varying only their size, and how wide and how frequent are the variations from the standard shape. A factor matrix which abolishes the general factor yields no means of answering such a question.

When we turn from the general factor of size or shape to the factors responsible for physical types, the factor pattern preferred by Thurstone again seems to leave the main question unanswered:

* With twins, however, the correlation for the bipolar factor (.87) appears to be higher than that for the general factor (.81).

just how important* are the tendencies towards distinct physical types as compared with the tendency for all men to preserve much the same general proportions? From his factor pattern (13, Table 6) it is impossible to say: and, since Thurstone's length factor and girth factor show little or no overlapping, for all we know, the physical types might be absolutely discrete classes. On the other hand, with the procedure I have suggested it is at once obvious that the type-factors contribute far less to the total variance than the general factor; moreover we can readily deduce a partial regression equation† for the bipolar factor, and so can give every man a measurement for his deviation towards the pachysomic direction or the leptosomic.

In this way we can examine the frequency distribution of the factor-measurements. We then discover that, so far from the types forming two distinct groups, the factor measurements are distributed in fairly close conformity with the normal curve—men who tend neither to a pachysomic nor to a leptosomic shape being by far the commonest (2). With the same procedure we can go on to correlate the factor measurements for physical type with factor measurements for temperamental type, e.g., the cyclothymic *versus* schizothymic. Here we have regularly found a positive but distinctly low correlation (usually about .20). And, if we wish, we can compare the assessments for body-build, reached in this way by a regression equation, with the assessments furnished by Sheldon's method of somatotypy, and thus decide which supplies the best basis for predictions.

6. *Simple Structure.* With this set of data, therefore, the divergences between Thurstone's results and my own seem clearly to be explained by his *a priori* determination to secure a unique "simple structure."‡ The criteria for such a structure (13, 156) prescribe

* Dearborn and Rothney, for example, believe they have no importance whatever. The notion of "types of body build," they say, "is a relic of a psychological error carried over into the field of physical development" (8, p. 290). Actually, however, their own table of correlations yields the same bipolar factor (7, 83).

† I fully agree with the objections urged by Dearborn and his colleagues against the various "indices of body build" in present use (8, 291-2).

‡ Since human bodies are tridimensional solids, the problem of measuring such solids in their simplest form, viz., as rectangular blocks, throws an instructive light on the issues arising in connection with body-build; and in another connection, Thurstone has shown how he would treat this simpler problem (14, 223-228). But here again in the factor pattern which he eventually reaches (l.c., Table 4) the most interesting feature about the boxes would escape us, namely, that in the collection actually measured there is an appreciable tendency for all the measurements to vary together; for "a box that is tall is likely to be thick and wide." Elsewhere Thurstone partly meets this criticism by introducing the concept of "second order factors": and in his most recent contributions seems to be approaching the view which I ventured on in 1909, namely, that the most revealing analysis of a correlation table may entail a general factor and a set of overlapping group factors.

that "each column of V should have at least r zeros" (V denoting the factor matrix and r the number of factors in it): i.e., after rotation, all the factors must be "group factors." The main reason for this is the desire to ensure that, so far as possible, the factorial composition of a trait remains invariant, even when the trait reappears in a different setting or "battery." I accept the aim; but I am doubtful about the method. Invariance, I believe, is best secured by seeking a "structure" which will include a general factor as well as overlapping group factors (unless, of course, such a general factor has been deliberately excluded by the mode of selecting traits or persons).

No doubt, before we attempt to interpret the nature even of the most general of our factors, we ought to plan an investigation so that the assessments included will indicate what this general factor is not, as well as what it is: in other words, the general factor should be deliberately converted into a group factor. For example, what some writers have taken to be an innate factor for general bodily growth might well be an environmental factor affecting nutrition during the years of growth: when, and only when, the general factor has virtually a zero saturation for nutritional assessments can we safely rule out that interpretation. The same holds good in determining a general factor obtained from psychological data: as I have pointed out elsewhere, until we have included it in a scheme measuring temperamental and moral qualities, we cannot safely assume that Spearman's g is a specifically cognitive factor, and then, of course, it becomes not a "general" but a "group" factor. But whenever we intend to seek a "simple structure" consisting solely of group factors, the inquiry should be expressly planned beforehand with that intention in view. If the data have not been collected with an eye to enforcing such a scheme, then it seems arbitrary to impose it, and highly precarious to reject a factor-pattern just because one of its columns contains no zeros. Indeed, so far from securing invariant factors, such a principle (in our experience at least) tends to alter them. Suppose, for instance, Hammond's anthropological colleague had been content with only one of the two Height measurements (say Standing Height), one of the two Chest Measurements (say Breadth), and one of the Head measurements (say Height) keeping the rest of the traits as before (a scheme of measurements far more like the usual anthropologist's list): then Thurstone's present factor-pattern would conform with my own, for (as will be seen from his table) it would exhibit one general factor and three group factors. But this would not constitute a "simple structure"; and simple structure would then have to be achieved (and could be achieved) by a different and indeed a simpler rotation. This, however, would yield different factors, not the same.

We cannot, as it seems to me, satisfactorily ensure or demonstrate the approximate identity of factors by merely procuring a "simple structure" for each isolated correlation-matrix as it is obtained. What we require is not a "simple structure" for this or that particular set of traits, but a "simple structure," or rather the simplest possible structure, for *all* available matrices in which a given group of traits occurs. And this in turn entails a formal test for the constancy of factors along the lines described above.

Of course, when our interest lies merely in determining the specialized group-factors, and we no longer want to establish the general factor or to assess its

relative importance, it will often prove more convenient to work with a factor-matrix possessing simple structure; but in that case we must deliberately select either traits or persons in such a way that no general factor is likely to intrude. Of the two alternatives, it is better, as it seems to me, to select persons rather than traits: for such a purpose, therefore, we should take persons of the same body size, or of the same general intelligence, or the same degree of general emotionality (cf. 5, 161), or whatever the field may be. But if we jump to this stage with no preliminary search for a possible general factor, then, I venture to think, we may be led into serious error. For example, some of the early attempts at determining imagery types overlooked the existence of a general factor for vividness in imagery of all concrete kinds; and in devising methods for occupational selection both in industry and in the fighting services we have continually found that performances in tests suggested for assessing specialized "factors" turn out to be dependent mainly on differences in "general intelligence."

Let me add, however, that I do not wish to deny that Thurstone's complete procedure—first, rotating to obtain a simple structure of correlated factors (with no general factor) and then factorizing the factors—may be of great interest as a means of procuring further information about the interrelations of the data. I merely consider that, for general descriptive purposes, a matrix containing general and group factors, secured by a more direct approach, is more appropriate and suggestive.

Conclusions

Let us now turn to the results so far obtained in the field of physical measurements by adopting the procedure described.

1. *Constancy of Factor Pattern*: (*a*) *with Different Ages.* On factorizing data from London school children we have found at almost every age much the same factors as with adults [cf. (5) and refs.]. But the largest batch of data that we have so far collected consists of measurements for 2,400 recruits for the Royal Air Force. The data were procured during the war and have only just been released for publication: the detailed correlation tables and factor saturations will be found in a paper by my assistant, Dr. Banks, and will shortly be published (2). The age of the men ranged from 17 to 38 years. But the total number is so large that it is possible to subdivide the entire batch up into eight age-groups, containing (with the exception of the oldest) well over 100 men in each. This enables us to study the stability of the factor-pattern from youth to middle age.

So long as the traits measured remain precisely the same in every group, the best method of comparison is that based on the original summational factors obtained direct from the observed correlations (preferably weighted). On comparing the saturations so computed, the factor-pattern proves to be remarkably constant. The correlations between the saturations for the general factor average .936; for the first bipolar, .975; for the second bipolar, .925. Except for the youngest age of all (17 years), none of the correlations is below .900.

TABLE 3

Correlations between Physical Measurements: 528 R.A.F. Men

	Height	Foot	Sitting Height	Arm	Leg	Thigh Length	Weight	Shoulder	Chest	Waist	Neck	Seat	Thigh Girth	Knee	Calf	Small	Wrist
Height	(.992)	.700	.766	.675	.853	.820	.636	.463	.331	.320	.276	.531	.305	.607	.300	.394	.383
Foot	.700	(.549)	.495	.595	.636	.606	.557	.339	.333	.310	.309	.494	.311	.604	.373	.426	.421
Sitting Ht.	.766	.495	(.874)	.380	.476	.432	.544	.368	.267	.246	.272	.495	.329	.466	.294	.376	.405
Arm	.675	.595	.380	(.618)	.664	.671	.516	.415	.371	.368	.258	.408	.266	.488	.257	.334	.228
Leg	.853	.636	.476	.664	(.809)	.832	.532	.365	.311	.329	.230	.450	.240	.556	.243	.364	.312
Thigh (L.)	.820	.606	.432	.671	.832	(.784)	.602	.431	.388	.409	.282	.533	.327	.606	.283	.403	.371
Weight	.636	.557	.544	.516	.532	.602	(.991)	.665	.794	.812	.737	.919	.866	.833	.745	.798	.705
Shoulder	.463	.339	.368	.415	.365	.431	.665	(.478)	.599	.530	.514	.595	.532	.518	.464	.452	.495
Chest	.331	.333	.267	.371	.311	.388	.794	.599	(.791)	.785	.661	.698	.698	.570	.558	.575	.564
Waist	.320	.310	.246	.368	.329	.409	.812	.530	.785	(.783)	.668	.775	.716	.593	.556	.618	.532
Neck	.276	.309	.272	.258	.230	.282	.737	.514	.661	.668	(.636)	.679	.686	.589	.566	.581	.596
Seat	.531	.494	.495	.408	.450	.533	.919	.595	.698	.775	.679	(.829)	.842	.778	.656	.746	.668
Thigh (G.)	.305	.311	.329	.266	.240	.327	.866	.532	.698	.716	.686	.842	(.839)	.716	.766	.761	.653
Knee	.607	.604	.466	.488	.556	.606	.833	.518	.570	.593	.589	.778	.716	(.841)	.681	.799	.673
Calf	.300	.373	.294	.257	.243	.283	.745	.464	.558	.556	.566	.656	.766	.681	(.727)	.779	.584
Small	.394	.426	.376	.334	.364	.403	.798	.452	.575	.618	.581	.746	.761	.799	.779	(.815)	.612
Wrist	.383	.421	.405	.228	.312	.371	.705	.495	.564	.532	.596	.668	.653	.673	.584	.612	(.568)

2. *Constancy of Factor Pattern*: (*b*) *with Different Nationalities.* Prior to this, and with the aid of other research-students (Cohen, Gamble, Eysenck), we had attempted factorial analyses for measurements reported from different national groups. The results have been summarized in a previous paper. There I pointed out that unexpectedly similar factor patterns have emerged for English, Irish, Welsh, Americans, and Jews [cf. (7)]. In these researches the data have been collected by different investigators, and the batteries of traits vary from one group to another. In such cases the factors furnished by the initial summational analysis are apt to differ somewhat; and the most stable results seem unquestionably to be secured by an analysis into general-plus-group factors.

Prof. Thurstone, as we have seen, argues that, where the batteries differ, it is wiser to judge the invariance of the factors from the "simple structure" obtained by a graphic rotation. On trying this procedure we find that, with free rotation, particularly when oblique factors are allowed, it is all too easy to produce similar patterns with different data. Moreover, the main changes in factor pattern arise, not so much from the inclusion of traits differently saturated with the constituent factors, but rather from including different *numbers* of traits representing the several groups. To compensate for such inequalities of number, the easiest procedure is to average the correlations for the several subgroups, or more simply to repeat the figures for certain traits until the numbers in each subgroup have been equalized. However, to our surprise, we find that a striking measure of agreement is exhibited, even in the results of an ordinary summational or centroid analysis.

3. *Specimen Analysis.* To illustrate the type of factor-pattern generally obtained, may I cite the results obtained with the largest and most recent series of body measurements we have so far examined? This consists of 17 body measurements for 528 men in the Royal Air Force.

The correlations are given in Table 3; and the factor saturations obtained by simple summation and by group factor analysis, respectively, are shown in Table 4. The group factor saturations are of special interest because they illustrate with unusual clarity a type of factor-matrix that we have also observed quite frequently in the analysis of mental measurements, namely, a pattern in which the entire set of traits yields first a pair of broad group factors, and then two pairs of narrower group factors, one pair in each of the broader groups.*

* The same type of pattern was also obtained on applying group factor methods to the data secured by another of our former students, Dr. Eysenck: here

(a) *The General Factor.* In this investigation as in all the preceding, the most conspicuous feature is the appearance of a general factor possessing positive saturations throughout: with summational analysis and an unselected population it contributes about 50 per cent of the total variance, ie., roughly as much as it would with mental or cognitive tests. This figure definitely disproves the view maintained by Spearman (and apparently accepted by Thomson and others) that there is some fundamental "contrast" between physical and mental measurements, and that, while mental measurements "tend to show a hierarchical order," physical measurements do not: or, to quote Thomson's way of putting it, whereas "the mind is so comparatively structureless, . . . the physical body has an obvious structure."

(b) *The Type-Factor.* Secondly, here as elsewhere the first bipolar factor displays a clear dichotomy between longitudinal traits (height, length of trunk, length of arms, legs, hands, etc.) and transverse or circumferential traits (breadth, depth, and girth of the trunk at various levels—shoulders, chest, waist, and hips—and, when they are included, girth of the various limbs). This result seems finally to confirm the distinction of body-builds into two main types (or rather tendencies), such as have been envisaged by so many writers; and further shows why this particular contrast is more conspicuous than any other. As a rule, the bipolar factor responsible for this subdivision contributes about 13 per cent to the total variance, whereas no other contributes more than 8 or 9 per cent.

These two factors have appeared in every age-group and in every national group so far examined. But when appropriate additions are made to the battery, we find that both the two main types tend to split up still further into subtypes and sub-subtypes. Thus, the leptosomic type can be subdivided into those who owe their longitudinal characteristics mainly to their long limbs, and those who owe them rather to their long trunk—a division suggestive of the classification into macroskelic and microskelic types proposed by Manouvrier and other anthropologists. Moreover, the macroskelic subtype in its turn seems further divisible into those whose limbs have relatively long upper segments and those whose limbs have relatively long lower segments, respectively. Again, the pachysomic or pyknic type can similarly be subdivided into two subtypes. There is first a subtype which owes its pyknic character mainly to lateral bony growth: this is shown chiefly by large measurements for breadth of trunk (e.g., at

small group factors for limb-length and trunk-length appear within the leptosomic group of traits, and for hip-girth, chest-girth, and chest-breadth within the pachysomic group.

TABLE 4
Analysis of Physical Measurements: 528 R.A.F. Men

Traits	A. Summational Factors				B. Group Factors						
	I	II	III	IV	Size	Lepto-somic	Pachy-somic	Trunk Length	Limb Length	Trunk Girth	Limb Girth
Height	.748	.648	−.061	−.101	.687	.665	–	.295	–	–	–
Foot	.644	.351	−.064	.087	.648	.384	–	.062	–	–	–
Sitting Ht.	.598	.373	−.374	−.488	.593	.175	–	.783	–	–	–
Arm	.601	.426	.219	.169	.581	.466	–	–	.098	–	–
Leg	.656	.585	.108	.159	.598	.669	–	–	.162	–	–
Thigh (L)	.702	.483	.172	.167	.691	.454	–	–	.558	–	–
Weight	.980	−.168	.021	−.053	.714	.202	.662	–	–	.108	–
Shoulder	.658	−.072	.148	−.135	.605	–	.264	–	–	.283	–
Chest	.743	−.329	.342	−.118	.528	–	.553	–	–	.475	–
Waist	.748	−.341	.321	−.067	.521	–	.607	–	–	.309	–
Neck	.683	−.378	.109	−.122	.457	–	.621	–	–	.165	–
Seat	.887	−.203	−.013	−.033	.451	–	.542	–	–	–	.086
Thigh (G)	.788	−.461	−.076	−.022	.479	–	.756	–	–	–	.323
Knee	.873	−.042	−.200	.193	.873	–	.217	–	–	–	.374
Calf	.706	−.363	−.245	.192	.483	–	.557	–	–	–	.425
Small	.786	−.283	−.256	.229	.614	–	.454	–	–	–	.470
Wrist	.707	−.226	−.151	−.057	.527	–	.458	–	–	–	.147
Sum of Squares	9.364	2.356	.684	.520	6.728	1.532	3.207	.704	3.48	.544	.675
Contribution to Variance	55.1	13.9	4.0	3.1	39.6	9.0	18.9	4.1	2.1	3.2	4.0

the shoulders and hips) and (when not obscured by fatty tissue) for circumference of trunk, particularly round the chest rather than the abdomen. There is a second subtype which owes its pyknic character rather to the overgrowth of soft tissues, and manifests itself in well-rounded trunk and limbs: this group is suggestive of Davenport's so-called "fleshy" type. It seems further divisible into a muscular subgroup and an adipose subgroup. Finally, as we have already seen with Hammond's data, there are minor factors for the growth of the extremities and for the shape of the head, which tend to show (as might be anticipated from observations on pathological cases) a marked independence of other factors.*

But all the factors so described are in the first instance merely statistical factors. To establish any interpretation in terms of causal agencies would require a separate type of investigation—researches (like those carried out by Julian Huxley in this country and by Schmalhausen and his collaborators abroad) upon the isometric and allometric factors underlying physical growth (cf, 6, 85 and refs.). We ourselves have attempted to factorize rates of growth in different traits, deduced from annual measurements of the same school children. The results suggest that the final physical type is in the more extreme cases a result of the accelerated or retarded growth rates (in longitudinal or transverse traits, respectively) which appear during the later phases of 'Streckung' and 'Füllung.' Parallel inquiries into analogous factors underlying mental growth, and into the relation of such tendencies to mental and temperamental development, are urgently needed; and, in our view, the methods of factorial analysis outlined here and in previous articles should provide a simple and useful technique.

Acknowledgments

For the measurements obtained from men in the Royal Air Force I am indebted to my colleague, Dr. Morant, who also supplied the correlations. For help with the factorizations I am further indebted to

* When the analysis aims at establishing a number of supplementary factors, it becomes important to have some method for estimating the significance of the factor saturations. For deciding when to stop factorizing it is sufficient to regard the saturation coefficient as an ordinary coefficient of correlation; and ascertain the probable or standard error from the usual tables. But this gives a value undoubtedly too low for testing significance. A simple but severer criterion is given by

$$\sigma = \frac{1 - r^4}{r\sqrt{Nd}}, \quad \text{where} \quad d = \frac{1}{2}\left(n - s - \frac{n+s}{n-s}\right),$$

and r denotes the saturation coefficient, N the number of persons measured, n the number of traits, and s the number of complete factors already extracted. Thus in Table 4a only four saturations for factor IV are definitely significant.

my assistant, Dr. C. Banks. We have to record our gratitude to Air-Marshall Sir Harold Whittingham, Director of Medical Services, R.A.F., for his willingness to facilitate the inquiries and to allow us to make suggestions in regard to the scheme of traits to be measured.

Note added June, 1947

A friendly critic asks for fuller details of previous work. A historical review of factorial studies of physical measurements will be found in Dr. Banks' thesis (1) which, I hope, may shortly be printed. In a paper just published (*Proc. Roy. Soc. Med.*, 1947, 40, 301-308). Dr. Tanner has summarized recent American work (including much that was inaccessible when the foregoing pages were written), and has discussed the relations between the conclusions drawn by American investigators and those I had deduced from British measurements.

In point of time the first to suggest analyzing a table of correlations between physical measurements, so as to reduce the data to terms of "hypothetical uncorrelated variables," was Karl Pearson (*Biometrika*, 1901, 1, 209). Pearson proposed using "the principal axes of the correlation ellipsoid" to represent what he called "ideal index-characters." But neither he nor his fellow-workers actually carried out any such calculations. His proposal, however, led to several fruitful suggestions for a more simple mode of analysis, since the table referred to by Pearson (with data from 3,000 adult criminals) lent itself admirably to an analysis into one general and two group factors, with no significant residuals. The advantage of Pearson's procedure is that it would permit a test for the significance of the factors, analogous to that employed in an analysis of variance: with the simpler analyses this procedure becomes of doubtful validity.

In their book on *Factor Analysis* (1941, p. 169), Holzinger and Harman quote an analysis of "8 physical variables for 305 fifteen-year-old girls" (taken from a Ph.D. dissertation by Frances Mullen, 1939, since published in *Child Development*, 11, 27f.). The full analysis "reveals a general size factor and two outstanding group factors identified with longitudinal and horizontal growth. After this solution was obtained . . . a solution with a general physical growth factor and a bipolar factor was then proposed." These results are in close accordance with the factor-patterns, showing a general and a bipolar factor (or alternatively one general and two group-factors), previously reported in 5 (pp. 184-5, summarized from an earlier L.C.C. Report: cf. also Burt, C., *The Backward Child*, 1937, pp. 135f. and p. 643).

In general the factor pattern obtained with what we term "the group factor method" (cf. Tables 1B and 4B) is very similar to that assumed by Holzinger's "bifactor method." But the working procedures are different. Prof. Holzinger's procedure is an extension of Spearman's product-formula; ours is an extension of the summation-formula: it is in fact merely an application to physical measurements of the method used for analyzing correlation-tables for mental tests in terms of a general, several group, and n specific factors (sometimes called the "three factor method," to distinguish it from Spearman's "two factor method": cf. 6, 140, 296 and refs.). I hope, however, to take up this point more fully in an early issue of the forthcoming *British Journal of Psychology: Statistical Section.*

REFERENCES

1. Banks, C. *Factor analysis applied to current problems based on data from H. M. Forces, 1945.* Unpublished Thesis: University of London Library.
2. Banks, C. and Burt, C. Factor analysis of body measurements for British adult males. *Annals Eugenics*, 1947, 13, 238-256.
3. Burt, C. The experimental study of general intelligence. *Child Study*, 1911, 4, 1-44, 92-101.
4. Burt, C. *Distribution and relations of educational abilities.* London: P. S. King, 1917.

5. Burt, C. The analysis of temperament. *Brit. J. med. Psych.*, 1938, 17, 155-88.
6. Burt, C. *The factors of the mind.* London: Univ. London Press, 1940.
7. Burt, C. The factorial study of physical types. *Man*, 1944, 44, 82-86.
8. Dearborn, W. F. and Rothney, J. W. M. *Predicting the child's development.* Cambridge, Mass.: Sci.-Art Publishers, 1941.
9. Eysenck, H. J. and Rees, W. L. A factorial study of some morphological and psychological aspects of human constitution. *J. Mental Science*, 1945, 91, 8-21.
10. Hammond, W. H. An application of Burt's multiple factor analysis to the delineation of physical types. *Man*, 1942, 42, 4-11.
11. Spearman, C. General intelligence objectively determined and measured. *Amer. J. Psych.*, 1905, 15, 201-93.
12. Thomson, G. The general factor fallacy in psychology. *Brit. J. Psych.*, 1919, 10, 319-26.
13. Thurstone, L. L. *The vectors of mind.* Chicago: Univ. Chicago Press, 1935.
14. Thurstone, L. L. Current issues in factor analysis. *Psych. Bull.*, 1940, 37, 189-236.
15. Thurstone, L. L. Factor analysis of body types. *Psychometrika*, 1946, 11, 15-21.

PSYCHOMETRIKA—VOL. 12, NO. 3
SEPTEMBER, 1947

MULTIPLE BI-SERIAL AND MULTIPLE POINT BI-SERIAL CORRELATION

ROBERT J. WHERRY
U. S. WAR DEPARTMENT

Normal equations, using data in various forms, are presented for securing the regression weights for prediction of a dichotomized criterion, and a simplified equation for the estimation of the multiple bi-serial or multiple point bi-serial, depending upon the proper assumption as to the nature of the distribution of the criterion, on the basis of these maximal weights is given also. The weights, unaffected by the assumption as to the nature of the criterion, are identical (or proportional) to those found by the discriminant function approach based upon analysis of variance. The author holds that the present multiple correlation approach is both easier and more informative than the discriminant function (analysis of variance) approach and suggests that the discriminant function be abandoned in favor of multiple bi-serial and/or multiple point bi-serial correlation and regression.

While multiple correlation, against a metricized criterion, is a commonly used procedure, no method for multiple bi-serial or multiple point bi-serial, involving a dichotomized criterion, appears in the literature. Instead we find the widespread belief that such procedures are illegitimate or impossible. The *Discriminant Function*, based upon the analysis of variance, is usually recommended as the proper approach. The discriminant function was first suggested by R. A. Fisher (1), and was introduced to psychologists by R. M. W. Travers (2) in this journal.

Later, H. E. Garrett (3) proved empirically that this discriminant function was "essentially a multiple regression equation." Actually both Travers and Garrett show the equivalence of weights obtained by the discriminant function approach to those obtained by the point bi-serial approach, but neither clearly brings out fully the relation of the method to correlation methods in general. Garrett is confusing when he claims that the discriminant function is applicable when "there is no independent criterion," and that the correlational (point bi-serial) approach is applicable when "an independent criterion is set up." Travers also seems to deny the presence of the criterion as such when he says that correlation (point bi-serial) will apply if members of each criterion group are "given a value of a dummy

variate." Actually the criterion is already present in all cases, and consists of the qualitative designation of class membership. The criterion is, however, unmetricized, and, when there are only two groups, the only type of correlation possible is either the bi-serial or point bi-serial, depending upon the most logical assumption as to the nature of the true distribution of criterion scores. The implication, by both Garrett and Travers, that the weights arrived at by the discriminant function are applicable only when the criterion consists of two discrete points rather than an actually metricizable continuum will be shown to be false.

The present paper will prove that the truly maximal weights (both Travers and Garrett refer to weights by the discriminant function as being only proportional to the true weights) can be obtained directly from correlational methods regardless of the assumption as to the nature of the criterion, since the weights are not affected by that assumption.

Since the discriminant function is based upon the analysis of variance approach which is both less informative (gives weights but not degree of relationship) and more burdensome (requires computation of variances separately for each group), the present author would suggest that the term and its accompanying methodology be rejected in favor of the easier and better understood correlation approach, and that the weights be called simply regression weights, and that their effectiveness be judged by the multiple bi-serial or multiple point bi-serial correlation resulting from their use in prediction.

Let us consider first the equations for the first-order bi-serial r,

$$r_b = \frac{M_p - M_q}{\sigma_T} \cdot \frac{pq}{y}, \tag{1}$$

and for the first-order point bi-serial r,

$$r_p = \frac{M_p - M_q}{\sigma_T} \cdot \sqrt{pq}. \tag{2}$$

It will be noted that these equations differ only in the last multiplier term and can be written more generally in the form

$$r_{b \text{ or } p} = \frac{M_p - M_q}{\sigma_T} \cdot K. \tag{3}$$

Now let us suppose that we desire to determine a composite score based upon several variables, so that

$$X_0 = W_1X_1 + W_2X_2 + \cdots + W_mX_m, \tag{4}$$

and desire to know the correlation of this composite score with some dichotomous criterion. Equation (3) will then become

$$_{C}r_{b \text{ or } p} = \frac{M_{Cp} - M_{Cq}}{\sigma_{C_T}} \cdot K. \qquad (3a)$$

But since

$$\sigma_{C_T} = \sqrt{\begin{array}{l} W_1^2\sigma_1^2 + W_2^2\sigma_2^2 + \cdots + W_m^2\sigma_m^2 + \\ 2W_1W_2\sigma_1\sigma_2r_{12} + \cdots + 2W_{m-1}W_m\sigma_{m-1}\sigma_m r_{(m-1)m} \end{array}}, \qquad (5)$$

$$M_{Cp} = W_1M_{1p} + W_2M_{2p} + \cdots + W_mM_{mp}, \qquad (6)$$

and

$$M_{Cq} = W_1M_{1q} + W_2M_{2q} + \cdots + W_mM_{mq}, \qquad (7)$$

if we substitute (5), (6), and (7) in (3a), we have

$$_{C}r_{b \text{ or } p} = \frac{W_1(M_{1p} - M_{1q}) + W_2(M_{2p} - M_{2q}) + \cdots + W_m(M_{mp} - M_{mq})}{\sqrt{\begin{array}{l} W_1^2\sigma_1^2 + W_2^2\sigma_2^2 + \cdots + W_m^2\sigma_m^2 + \\ 2W_1W_2\sigma_1\sigma_2r_{12} + \cdots + 2W_{m-1}W_m\sigma_{m-1}\sigma_m r_{(m-1)m} \end{array}}} \cdot K. \qquad (3b)$$

If we further desire that the weights actually maximize $_{C}r_{b \text{ or } p}$ so that the composite correlation will become a multiple correlation, we can achieve this indirectly by maximizing the logarithm of $_{C}r_{b \text{ or } p}$. We first must substitute $D_i = M_{ip} - M_{iq}$ for each difference term in the denominator of (3b), to obtain

$$_{C}r_{b \text{ or } p} = \frac{W_1D_1 + W_2D_2 + W_3D_3 + \cdots + W_mD_m}{\sqrt{\begin{array}{l} W_1^2\sigma_1^2 + W_2^2\sigma_2^2 + \cdots + W_m^2\sigma_m^2 + \\ 2W_1W_2\sigma_1\sigma_2r_{12} + \cdots + 2W_{m-1}W_m\sigma_{m-1}\sigma_m r_{(m-1)m} \end{array}}}. \qquad (3c)$$

Then, letting

$$A = W_1D_1 + W_2D_2 + W_3D_3 + \cdots + W_mD_m \qquad (8)$$

and

$$B = W_1^2\sigma_1^2 + W_2^2\sigma_2^2 + \cdots + W_m^2\sigma_m^2 + 2W_1W_2\sigma_1\sigma_2r_{12} + \cdots + 2W_{m-1}W_m\sigma_{m-1}\sigma_m r_{(m-1)m}, \qquad (9)$$

we can write

$$_{C}r_{b \text{ or } p} = \frac{A}{\sqrt{B}} \cdot K. \qquad (3d)$$

Taking the logarithms of both sides of (3d) gives

$$\log r = \log A + \log K - \tfrac{1}{2}\log B. \qquad (3e)$$

To maximize $\log R$ (and hence R) we take the partial derivatives of $\log R$ with respect to each value of W_i and set these equal to zero; thus:

$$\frac{\delta \log R}{\delta W_i} = \frac{\delta A/\delta W_i}{A} + \frac{\delta K/\delta W_i}{K} - \frac{1}{2}\frac{\delta B/\delta W_i}{B} = 0\,, \tag{10}$$

But, since the partial of K with respect to W_i is zero since K is independent of the W values, (10) reduces to

$$\frac{\delta \log r}{\delta W_i} = \frac{\delta A/\delta W_i}{A} - \frac{1}{2}\frac{\delta B/\delta W_i}{B} = 0. \tag{11}$$

This tells us that the normal equations for the weights are independent of K, i.e., are independent of our assumption concerning the nature of the criterion. From (11) we can write

$$\begin{aligned}
\frac{\delta \log r}{\delta W_1} &= \frac{D_1}{A} - \frac{1}{2}\frac{(2W_1\sigma_1^2 + 2W_2\sigma_1\sigma_2 r_{12} + \cdots + 2W_m\sigma_1\sigma_m r_{1m})}{B} = 0 \\
\frac{\delta \log r}{\delta W_2} &= \frac{D_2}{A} - \frac{1}{2}\frac{(2W_1\sigma_1\sigma_2 r_{12} + 2W_2\sigma_2^2 + \cdots + 2W_m\sigma_2\sigma_m r_{2m})}{B} = 0 \qquad (12) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\frac{\delta \log r}{\delta W_m} &= \frac{D_m}{A} - \frac{1}{2}\frac{(2W_1\sigma_1\sigma_m r_{1m} + 2W_2\sigma_2\sigma_m r_{2m} + \cdots + 2W_m\sigma_m^2)}{B} = 0.
\end{aligned}$$

The equations above can be simplified to obtain

$$\begin{aligned}
W_1\sigma_1^2 + W_2\sigma_1\sigma_2 r_{12} + \cdots + W_m\sigma_1\sigma_m r_{1m} &= D_1\, B/A \\
W_1\sigma_1\sigma_2 r_{12} + W_2\sigma_2^2 + \cdots + W_m\sigma_2\sigma_m r_{2m} &= D_2\, B/A \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \qquad (13) \\
W_1\sigma_1\sigma_m r_{1m} + W_2\sigma_2\sigma_m r_{2m} + \cdots + W_m\sigma_m^2 &= D_m\, B/A\,.
\end{aligned}$$

If we temporarily disregard the multiplier B/A we obtain

$$\begin{aligned}
W_1\sigma_1^2 + W_2\sigma_1\sigma_2 r_{12} + \cdots + W_m\sigma_1\sigma_m r_{1m} &= D_1 \\
W_1\sigma_1\sigma_2 r_{12} + W_2\sigma_2^2 + \cdots + W_m\sigma_2\sigma_m r_{2m} &= D_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \qquad (14) \\
W_1\sigma_1\sigma_m r_{1m} + W_2\sigma_2\sigma_m r_{2m} + \cdots + W_m\sigma_m^2 &= D_m\,.
\end{aligned}$$

Since the number of W_i unknown values is equal to the number of equations in set (14), we can exactly satisfy these equations. If the equations in set (14) are substituted in equation (8) we find that

$$\begin{aligned}
A = {} & W_1(W_1\sigma_1^2 + W_2\sigma_1\sigma_2 r_{12} + \cdots + W_m\sigma_1\sigma_m r_{1m}) \\
& + W_2(W_1\sigma_1\sigma_2 r_{12} + W_2\sigma_2^2 + \cdots + W_m\sigma_2\sigma_m r_{2m}) \qquad (8a) \\
& + \cdots + W_m(W_1\sigma_1\sigma_m r_{1m} + W_2\sigma_2\sigma_m r_{2m} + \cdots + W_m\sigma_m^2),
\end{aligned}$$

whence by expanding and collecting like terms we have

$$A = W_1^2\sigma_1^2 + W_2^2\sigma_2^2 + \cdots + W_m^2\sigma_m^2 + 2W_1W_2\sigma_1\sigma_2 r_{12} + \cdots + 2W_{m-1}W_m\sigma_{m-1}\sigma_m r_{(m-1)m} . \tag{8c}$$

But (8c) is now identical with (9), whence

$$A = B . \tag{15}$$

This finding has two important consequences. It means first that the ratio B/A disregarded in going from set (13) to set (14) was actually equal to unity and so could be legitimately disregarded as a multiplier. In the second place we can rewrite (3d) as

$$\begin{aligned} {}_cR_{b \text{ or } p} &= \sqrt{A}\cdot K \\ &= \sqrt{W_1D_1 + W_2D_2 + \cdots + W_mD_m}\cdot K , \end{aligned} \tag{3f}$$

which will greatly facilitate the estimation of ${}_cR_{b \text{ or } p}$ after the weights have been established, by solving the equations in set (14) by the Doolittle or any other economical method.

While set (14) permits an exact solution of the weights, W_i, the equations are not in their most useful form, since such terms as $\sigma_i\sigma_j r_{ij}$ are difficult to compute. Several variants of (14) are possible, however. If we multiply each equation in set (14) by N and make the substitutions

$$N\sigma_i\sigma_j r_{ij} = \sum x_i x_j \quad \text{and} \quad N\sigma_i^2 = \sum x_i^2 , \tag{16}$$

we obtain

$$\begin{aligned} W_1\textstyle\sum x_1^2 + W_2\sum x_1x_2 + \cdots + W_m\sum x_1x_m &= ND_1 \\ W_1\textstyle\sum x_1x_2 + W_2\sum x_2^2 + \cdots + W_m\sum x_2x_m &= ND_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ W_1\textstyle\sum x_1x_m + W_2\sum x_2x_m + \cdots + W_m\sum x_m^2 &= ND_m . \end{aligned} \tag{14a}$$

Or, again multiplying each equation in set (14a) by N and making the substitutions

$$N\textstyle\sum x_i x_j = N\sum X_iX_j - \sum X_i\sum X_j = L_{ij}$$

and

$$N\textstyle\sum x_i^2 = N\sum X_i^2 - (\sum X_i)^2 = L_{ii} , \tag{17}$$

we obtain

$$\begin{aligned} W_1L_{11} + W_2L_{12} + W_3L_{13} + \cdots + W_mL_{1m} &= N^2D_1 \\ W_1L_{12} + W_2L_{22} + W_3L_{23} + \cdots + W_mL_{2m} &= N^2D_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ W_1L_{1m} + W_2L_{2m} + W_3L_{3m} + \cdots + W_mL_{mm} &= N^2D_m . \end{aligned} \tag{14b}$$

Or, if we return to set (14) and multiply throughout by K/σ_i, where the i is the subscript of the D of the row in question and K is the appropriate multiplier depending upon the assumption as to the nature of the criterion, and we make the substitutions of

$$W_i' = W_i \sigma_i K, \tag{18}$$

and

$$r_{i(b \text{ or } p)} = D_i K/\sigma_i, \tag{3a}$$

we obtain

$$\begin{aligned}
&W_1' + W_2' r_{12} + W_3' r_{13} + \cdots + W_m' r_{1m} = r_{1 (b \text{ or } p)} \\
&W_1' r_{12} + W_2' + W_3' r_{23} + \cdots + W_m' r_{2m} = r_{2 (b \text{ or } p)} \\
&\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&W_1' r_{1m} + W_2' r_{2m} + W_3' r_{3m} + \cdots + W_m' = r_{m (b \text{ or } p)}.
\end{aligned} \tag{14c}$$

By combining (18) and (3f), we can write

$${}_cR_{b \text{ or } p} = \sqrt{\left(\frac{W_1'}{\sigma_1} D_1 + \frac{W_2'}{\sigma_2} D_2 + \cdots + \frac{W_m'}{\sigma_m} D_m \right) \cdot K}, \tag{3g}$$

and the W_i values can be obtained from the W_i' values by

$$W_i = W_i'/(\sigma_i K). \tag{18a}$$

The W_i values can then be obtained by the use of any one of the sets of equations (14), (14a), (14b), or (14c) depending upon whichever values are most easily obtainable. It should perhaps be noted that Travers used set (14a) (nearly) while Garrett used set (14c). If the inter-r values are already available or if test selection procedures are to be used, set (14c) seems indicated. If the inter-r values are not available and if one does not desire to use test selection, the use of set (14b) is probably most economical, since only the numerators of the r's will have to be obtained and no transformation (from W_i' to W_i) will be necessary. If the numbers involved in set (14b) are too large for easy handling, the normal equations can be reduced to set (14a) by dividing all values by N, or can be reduced to set (14) by dividing all numbers through by N^2, without in any way affecting the values of W_i resulting from their solution by the Doolittle method or any other economical method for solving simultaneous equations. After the W_i values are obtained, the multiple bi-serial or multiple point bi-serial coefficient of correlation can be obtained by substitution in (3f). In estimating the values of this R in future populations, the standard correction for degrees of freedom used up by m variables would apply.

REFERENCES

1. Fisher, R. A. The use of multiple measurements in taxonomic problems, *Annals Eugenics*, 1937, 7, Part II, 179-188.
2. Travers, R. M. W. The use of a discriminant function in the treatment of psychological group differences. *Psychometrika*, 1939, 4, 25-32.
3. Garrett, H. E. The discriminant function and its use in psychology. *Psychometrika*, 1943, 8, 65-79.

PSYCHOMETRIKA—VOL. 12, NO. 3
SEPTEMBER, 1947

CONFIRMATION AND CLARIFICATION OF PRIMARY PERSONALITY FACTORS

RAYMOND B. CATTELL
UNIVERSITY OF ILLINOIS

In connection with a study bridging rating, questionnaire, and objective test factors, confirmation was sought with respect to the twelve personality factors previously found for young adult men. Variables were chosen to clarify and discriminate the nature of related factors. Ratings of and by 373 students were obtained, and the present study describes the separate factorization for the 133 men among them. Factorization yielded eleven factors, of which, on "blind" rotation for simple structure, 9 or 10 proved to be identical with those of the previous study. A new factor *M* is described.

1. *Perspective on Plan of Integrating Research*

This study is one of a coordinated trio directed to (a) clearing up the number and nature of personality factors in three fields: 1. Behavior rating data (behavior in real life situations), 2. Questionnaire data, and 3. Objective test data; (b) determining the relationships among the factors in these three fields, to test the hypothesis that they are actually the *same* factors outcropping in different media.

This hypothesis supposes that the factors so far discovered (3) are in some sense functionally unitary personality traits, which manifest themselves alike in the total behavior (Behavior ratings or *BR* factors), in laboratory test situations (Objective tests or *T* factors) and in questionnaire self ratings (*Q* factors). The *Q* factors may be considered "mental interiors" — the view of the external behavior (*BR* and *T*) factor patterns as seen, or rather "felt," by the subject. The bridging of *BR*, *Q*, and *T* factors, which has not previously been attempted*, demands that the same subjects be used for all three series of testings, a requirement which, because of its exactions in time, is not easily attained. In the present research we started with 400 student† subjects for the *BR* and *Q* studies, who dwindled somewhat

* There have, of course, been numerous studies bridging in terms of simple *BR*, *Q*, and *T variables*, but not in terms of matching definite unitary traits.

† These were students in a state university, taking a rather large section of the population, swollen still further by the first returning veterans; so that the study is, fortunately, not open to the usual criticism of the use of student subjects, namely that they represent an extremely small and biased sample of the general population. Certainly, in spite of the comments of teachers, these subjects are somewhat selected in intelligence and probably to a lesser extent in character; but not enough to do more than alter the relative variance due to different factors.

before the *T* studies were finally completed. The present article describes the *BR* factor research only, the remaining researches being in press elsewhere.

2. *The Primacy of Behavior Rating Studies*

The behavior rating study is considered the keystone of the present research design. It is of no avail to establish factors in questionnaires or laboratory test variables unless such factors are related to the total life situation and the gross behavior patterns known to be of importance in every-day life. The choice of variables in this latter realm can be made truly representative, through use of the *personality sphere* concept*, whereas the choice of variables in tests may lead to general factors which are anything but general factors in the total behavior, occupying, indeed, some extremely narrow segment.

Some psychologists have regarded the use of ratings as suspect, and rightly so in the uncontrolled fields of applied psychology in which they have been most used. But no apology is necessary for their use here, because (a) As indicated above, only ratings on total behavior *in situ* will give us the true frame of reference we desire as a foundation for studying the total personality. (b) We propose, as soon as the rough scaffolding of factors in ratings is obtained, to use this as a means to set up and proceed to factors in objective tests measuring these same factors. (c) Granted proper planning of the rating situation, the conventional objections to ratings, i.e., assessments of behavior in real life situations, in the writer's opinion, no longer apply with any force.

The proper planning of rating, as discussed elsewhere (1, 4, 8), requires such essentials as that the raters live with the ratees; have no official relation to them; are not asked to rate more than about twenty people, because more cannot be known well; deal with all subjects at one time with respect to one trait and not all traits with respect to one subject; have traits behaviorally defined and illustrated; have the ratings pooled for at least ten judges to cancel individual perspectives.

When these are faithfully observed, the only remaining source of systematic error is halo or stereotype. This effect, though much

* For discussion of this concept see (3, p. 215), whence the following quotation: "The personality sphere may be considered to represent a complete surface, constituted by many small 'trait areas' . . . the whole constituting an endless but finite continuum of behavior meaning. . . It is part of our assumption, from the study of language, that no large area remains completely neglected by existing (trait name) vocabulary." The sphere has the space dimensionality of the principal factors required to describe personality.

talked about, has never been demonstrated in trained raters, and in even untrained raters the alleged halo is open to the alternative interpretation that desirable qualities really do go together. Sociologists have demonstrated an occasional stereotype (but not in personality, apart from superficial cultural, class, or national patterns); yet they have begged the question as to how the stereotype arises if it is not based on experience. The stereotype of schizophrenia is hardly one which the average man would invent. In our present research, we should in any case reply that though the popular mind might create one halo or stereotype it is rather far-fetched to suppose it would entertain twelve or more — the number of factors found in our researches. Further, our raters were not psychology students, but unsophisticated in typology. Even had they been psychology students, the sad experience of most examiners is that average undergraduates cannot collectively carry accurately in mind two or three, let alone a dozen, descriptions of commonly agreed personality types! Apart from these considerations the writer believes the evidence to show that rating (or, better, ranking by paired comparison) of all subjects on one trait at a time helps to break down any tendency that may exist for halo formation.

The conditions of rating here were that the men were taken in residential groups of 17 in each. Each person was rated, by everyone as *average, above average,* or *below average* in a given trait, with instruction to use these categories as nearly as possible in the general frequency 2:1:1 for that trait. In addition the rater was asked to star the very highest and the very lowest person in the group. Values 0, 1, 2, 3, 4 and 5 were given—i.e., 2 for average, 0 for the one extreme low case, 5 for the one extreme high, and the results were summed for the first eight and for the second eight raters. These two pooled halves were used as a basis for calculating the reliability coefficient, before being added to give a single score.

One important difference of practice exists between this and the previous similar research (1), and it was adopted, not because we consider it superior, but because it seemed desirable to see whether the same results would be obtained independently of rating technique. In the previous study (1), correlations were carried out within each small group (14 groups of 16 each) and then averaged via Fisher's z function. In the present study all the final ratings were put in a common population and then correlated. For it will be noted that though a general directive was given to approximate to a normal distribution in each group the rater was explicitly directed to take the average student as his average, whereby a group high or low as a whole could be scored accordingly.

As the results show, no untoward lack of flexibility seems to have existed in between-group variability of assigned scores. But we consider that error *would* have been introduced had we mixed the ratings of men by men with those of women by women. For example, the most aggressive woman might be no more aggressive than the average man. Further, such a mixture would have prevented or obscured the eventual emergence of any factor patterns which might be peculiar to one sex. Therefore, in spite of the introduction of some nicer problems of coordination which would occur when we should come to intercorrelating *BR*, *Q* and *T* factors later, we decided to keep the *BR* studies of men and women separate both in rating and factorization. This first paper describes the factorization for 133 men, since men were the subjects of the original personality factor study (1). The results for the 240 women are set out and compared elsewhere (5).

3. *Description of the Variables*

The success of factorial research depends on a wise choice of representative variables, even more than it does on a correct choice of representative populations. In the initial research in any field it is necessary to have an entirely unbiased sampling of possible variables, as defined in the personality sphere and as carried out in the main preceding study (1). Thereafter the choice depends on purpose. If the researcher is investigating one factor intensively he will pack variables closely in that area of behavior, as far as his insight and intuition permit, and represent the remaining factors each by the minimum number of known "landmark" variables (generally two) necessary to indicate their identity. In the present study our purpose was to clarify, and bring out more sharply, the nature of the factors already shown to be operative in the total personality sphere. We proceded, therefore, by the following rules:

(1) Each factor to be represented by a minimum of two variables, chosen from the six most highly saturated in each factor in the previous study.

(2) Each factor to be represented by a variable which, from previous studies of all kinds (3), seemed most likely to be the essence of the factor. That is, we attempted to find a purer measure, with higher saturations than any previously found. This is part of an iterative, "distillation" procedure, in which each research profits by the preceding to get closer to the nature of the factor.

(3) The definition of landmark variables to be re-shaped slightly from their original "logical trait" form to "organic trait unity" form*

* The variables with which an experimenter starts are normally defined by arbitrary, logical limits existing in the mind of the experimenter. In time the

in the light of associates discovered for them in the previous factorization.

That the objectives of (1) (2) and (3) were to some extent achieved is shown (since there is no increase of reliability of ratings *per se* in the present research to account for greater inter-correlation) by the fact that the communality for most variables was definitely increased in the present study.

(4) Where factors of a similar or cooperative nature existed in the previous study, e.g., the two schizoid factors *A* and *H*; the three character factors *C*, *G*, and *J*, to choose variables that would "pull them apart" better, i.e., bring out the finer characteristics by which they differ.

(5) To give relatively more variables to factors needing sharper definition, particularly the factors (*H* and onwards) of smaller variance (in the original study). Thus four variables were inserted to bring out factor *K*, but only three for the well known dominance factor *E* and only one — an actual intelligence test — to mark the "intelligence-in-personality" factor *B*.

The defined variable list given to the raters dispersed these factor groupings and reversed the polarity of items indiscriminately, as shown below.

DEFINITIONS*

1. *Readiness to cooperate* Generally tends to say yes when invited to cooperate. Outgoing. Ready to meet people at least halfway. Finds ways of cooperating despite difficulties.	vs	*Obstructiveness* Inclined to raise objections to a project, cynical or realistic. "Cannot be done." Uninterested or unfavorable attitude to joining in. Inclined to be "difficult."
2. *Emotionally Stable* Can be depended upon to look at questions objectively, without emotional prejudice, and in the same constant light from day to day. Above emotion in his judgments. Dependable and realistic.	vs	*Changeable* Sees things in terms of the emotion of the moment. Emotional bias changes from day to day and place to place. Does not remain the same person from day to day. Undependable.

* The items of behavior included under one definition are put together partly on logical grounds, but also, and predominantly, on the evidence of previous research on the constituents of correlation clusters or "surface traits." Thus, in trait 20, miserliness and orderliness are included, though not logically related, and in trait 13, proneness to emotional display of both joy and sorrow are together, because of previous evidence of the existence of a trait of "general emotionality." (See 3, Chapter 8, for this evidence).

---

form of the unitary trait as it exists in nature emerges. The experimenter can then measure variables in the same area, defined anew better to align themselves with the true, organic nature of the functional unity. The contrast of "logical" and "organic" trait unities can be readily illustrated by comparing the trait unities of the phrenologist with those found in later studies of brain function, or by the modern factor analysts. (See 3, 126).

3. *Attention-getting* Shows off in company. Not happy unless in center of the stage. Talks about self, accomplishments, important friends, etc. Likely to show some "affected" behavior.	vs	*Self-sufficient* Not under compulsion to impress or to get sympathy or attention.
4. *Assertive, Self-assured* Assumes he can impose his (or her) will on others. Tends to lead or influence his associates. Tends to dominate. Tends to be boastful and assertive. Not held back by doubts. Invulnerable self-esteem.	vs	*Submissive* Tends to let other people have their way. Tends to back down in a conflict. Humble, quiet, retiring. Not sure he is right. "Embarrassable."
5. *Depressed, Solemn* Earnest and solemn most of the time. Not easily moved to laughter. Seeming slow and depressed rather frequently.	vs	*Cheerful* Generally bubbling over with good cheer. Optimistic. Enthusiastic. Prone to cheerful, witty remarks. "Laughterful."
6. *Frivolous* Does not seem to take responsibilities seriously. Undependable. Thoughtless. Refuses to accept responsibilities of his age.	vs	*Responsible* Has a sense of responsibility to his parents, community, etc. Can be depended upon to be loyal to agreed standards. Trustworthy.
7. *Attentive to People* Interested in people, their troubles, their personalities. Makes friends with people and remembers their personal interests. Spends much time in dealing with people.	vs	*Cool, Aloof* Tends to be indifferent to, and to ignore, people. Gives the impression of brooding on his own thoughts or of being cold and indifferent.
8. *Easily Upset* Easily embarrassed or put off balance in conversation. Gets confused in emergency. Blushes, shows excitability, becomes incoherent. (*Not* general emotionality, but momentary "nervousness.")	vs	*Unshakable Poise, Tough* Self-possessed, hard. does not lose composure, e.g., through emotional provocation.
9. *Languid, Slow* Lacks vigor. Has an absent-minded manner. Vague and slow in speech, etc.	vs	*Energetic, Alert* Vigorous, quick, and alert most of the time. Spirited. Enterprising. Always "on the ball."
10. *Boorish* Rather ignorant. Unreflective. Does not read much or enjoy intellectual problems. Narrow, simple interests.	vs	*Intellectual, Cultured* Has wide interest and knowledge, especially in intellectual matters. Is thoughtful and introspective about life. Enjoys analytical, penetrating discussions in small groups.

11. *Suspicious* Believes rather too quickly that he is being unfairly treated. Imagines on insufficient grounds that certain people strongly dislike him. Inclined to brood over his troubles. Interprets things as having reference to himself when none is intended. Feels persecuted.	vs	*Trustful* Free from suspicion.
12. *Good-natured, Easygoing* Does not mind when people use his property, time, or energy. Generous, gives people "the benefit of the doubt" when their motives are in question. Warmhearted.	vs	*Spiteful, Grasping, Critical* Gets irritable, "awkward," or resentful if property or other rights are trespassed on. Inclined to be "close" and grasping. Is generally surly, hard, and spiteful.
13. *Calm, Phlegmatic* Remains relatively calm in dispute, danger, social hilarity, temptation. Shows few signs of emotional excitement. (Not merely controls impulses, but scarcely seems to have them.)	vs	*Emotional* Gets emotional (anger, fear, jollity, sex, sorrow, or disgust) on slight provocation. Frequently emotional and passionately excited. Shows marked signs of emotion, even if controlled.
14. *Hypochondriacal* Dwells on illnesses or hurts a good deal. Is afraid is going to die when has relatively trivial illness. Fusses a good deal over bodily symptoms.	vs	*Not so* Absence of *H*.
15. *Mild, Self-effacing* Gentle-tempered. Blames himself (or nobody) if things go wrong.	vs	*Self-willed, Egotistic* Goes his own way regardless of others. Blames others, not himself, whenever there is conflict or things go wrong. Headstrong. Predatory—tends to use other people for own ends.
16. *Silent, Introspective* Says very little; gives the impression of being introspective and occupied with thoughts.	vs	*Talkative* Talks a lot, to everybody.
17. *Persevering, Determined* Sees a job through in spite of difficulties or temptations. Strong-willed. Persisting in his motives. Painstaking and thorough.	vs	*Quitting, Fickle* Gives up rather easily. Led astray from main purposes by stray impulses. Slipshod—does not finish a job thoroughly.
18. *Cautious, Retiring, Timid* Avoids the strange and new. Looks at all aspects of a situation overcautiously. Keeps clear of difficulties. Uninquiring, lacking in desire to try new things.	vs	*Adventurous, Bold* Rushes in carefree fashion into new experiences, situations, emergencies. Ascendant: ready to meet anything. Happy-go-lucky. Has a great appetite for life.

19. *Hard, Stern* Toughly "realistic" about problems. In extreme instances cynical. Looks at questions in a cold, objective fashion. Unaffected by personal and emotional appeals.	vs	*Kindly, Soft-hearted* Tender-hearted. Cannot bear to see suffering unalleviated. In extreme instances sentimental and unrealistic. Of a grateful, understanding disposition.
20. *Insistently Orderly* Tidy, over-precise, especially over details. Drives other people to be the same. Strict, fussy, pedantic. Insists on everything being orderly. (In these respects rather "uncomfortable to live with.") Seems unable to relax. Miserly.	vs	*Relaxed, Indolent* Rather careless of detail. Lazy. Careless over expenditures. Has no difficulty in relaxing. Enjoys ease.
21. *Polished* Polite and charming in social situations. Deals with people gracefully and skillfully. Refined with speech, manner, etc. Familiar with good etiquette.	vs	*Clumsy, Awkward* Clumsy in social situations. Crude in speech, manner, etc.
22. *Prone to Jealousy* Becomes readily jealous of people. Unreasonably hostile.	vs	*Not Prone to Jealousy*
23. *Rigid* Always does things in one particular way. Is nonplussed if his routine of life is upset. Sticks to his own ideas and does not adapt to ways of doing things different from his own. Follows the letter of the law, not merely the spirit. Does not change and broaden with circumstances. Conservative in personal habit.	vs	*Adaptable* Adapts his habits and ways of thinking to those of the group. Is satisfied with compromises—does not stand meticulously on the letter of the law. Is not upset, surprised, baffled, irritable if things are different from what he expected.
24. *Demanding, Impatient* Expects a lot from other people. Makes constant demands. Self-centered. Forgets needs of others. Impatient.	vs	*Emotionally Mature* Self-sufficient. Capable of renunciation without fuss. Thinks of others.
25. *Unconventional, Eccentric* Breaks conventional rules. Does not mind being different from other people in dress, manners, interests. Has well-marked individual ways (of which he is aware) and even fads and eccentricities.	vs	*Conventional* Conforms to the accepted standards and ways. Seems distressed if he finds he is being "different."

26. *Placid* Calm, tough. "What's the fuss about?" attitude.	vs	*Worrying, Anxious* Worries constantly, sensitive, hurried; seems to suffer from anxieties without adequate cause; slight suppressed agitation much of the time.
27. *Conscientious* Careful about principles of conduct. Guided by ideals of truthfulness, honesty, unselfishness. Scrupulously upright where personal desires conflict with principle.	vs	*Somewhat Unscrupulous* Inclined to somewhat shady transactions. Not too careful about right and wrong where own wishes are concerned. Not particularly just, honest or unselfish.
28. *Composed* Quite free from shyness. Completely at home in company or with strangers.	vs	*Shy, Bashful* Not necessarily unsociable, but extremely shy and sensitive. Goes out of his way to avoid meeting people. Tense when speaking in a group. Tries to avoid occasions when he will be in limelight.
29. *Sensitively Imaginative* Inclined to be governed by a vivid imagination. Thinks of unusual angles and aspects of a question. Sensitive to a multitude of emotional and other possibilities not realized by the average person. Intuitive, more interested in mental than material and practical aspects of a situation.	vs	*Practical, Logical* Solves questions in a logical matter-of-fact fashion which often ignores fine points or unusual possibilities. Heavily and "blindly" logical, refusing to see intangibles. More interested in material than mental aspects of a situation.
30. *Neurotic Fatigue* Seems to get tired and overwrought. Is irrationally irritable. Jumps when spoken to. Shows facial tics and other signs of "nervousness" (e.g., fidgeting, tremor, digestive disturbances, poor memory). Constantly complains of fatigue.	vs	*Absence of Neurotic Fatigue*
31. *Esthetically Fastidious* Artistically sensitive in clothes, surroundings, art. Fastidious, not too easily pleased.	vs	*Lacking Artistic Feeling* Not showing taste in clothes, etc. Not interested in artistic subjects. Insensitive to esthetic effects.
32. *Marked Interest in Opposite Sex* Dates a good deal. Shows much interest in opposite sex.	vs	*Slight Interest in Opposite Sex* Dates very little. (And not merely from lack of opportunity).
33. *Frank, Expressive* Comes out readily with his real feelings on various questions, so that you know where you stand with him.	vs	*Secretive, Reserved* Keeps his thoughts and feelings to himself. Often leaves you puzzled as to the motives for his actions. Inscrutable.

Expresses his feelings, sad or gay, easily and constantly. Easy to understand.		Does not give away information for the fun of it.
34. *Gregarious, Sociable* Likes to be in large groups. Seeks people out for the sake of company. Likes parties as often as possible. Not fond of being alone.	vs	*Self-contained* Does not seem to miss company of others. Goes own way.
35. *Dependent, Immature* Emotionally and intellectually dependent on others. Generally adopts the opinion of the group or of authority without much thought. Thoughts vague and confused. Rather immature.	vs	*Independent-minded* Thinks things out for himself and adopts a clear and definite independent position. Tends to be a leader in discussion. Is interested in public opinion and how to shape it.

To the above variables we added an estimate of *36. Intelligence vs Deficiency of Intelligence* obtained by use of the A.C.E. test in the normal course of college entrance. This one objective test was common to both the behavior rating factorization and the objective test factorization.

A representative set of reliability coefficients obtained by correlating the pool of the first eight with that of the second eight raters, and correcting by the Spearman-Brown formula to the full 16 raters, is as follows: Variable (1) .58; (2) .64; (4) .69; (8) .54; (9) .53; (12) .50; (17) .68; (26) .60; (27) .60; (28) .74; (29) .52; (30) .51. These run a full twenty points lower than the figures for the corresponding women's group, which is entirely in accord with our observations of the relative conscientiousness with which ratings were carried out in the two groups. However, it is interesting to notice that the agreement between the two groups as to which variables are better grasped and rated is almost perfect. In both, common and obvious forms of behavior, such as "shyness," "boorishness," or "frivolousness" are most reliably related; then follow slightly interpretive traits like "cooperative" or "conventional"; then judgments of emotionality like "easily upset." "Energetic, alert" somewhat surprisingly falls low, and a very interpretive trait like "neurotic fatigue" lowest of all.

4. *Conduct of the Factor Analysis*

The correlation matrix was factored by the centroid method. The application of Tucker's criterion (9) suggested that extraction was complete after eleven factors had been removed. However, we planned to proceed by Thurstone's practice of extracting a probable excess of factors, deciding thereafter the true number by the verdict of how many could be reduced to residuals in the process of rotation. Thirteen factors were therefore initially abstracted, as shown in Table 1.

TABLE 1
Unrotated, Centroid Factor Matrix

Variables	Factors													
	1	2	3	4	5	6	7	8	9	10	11	12	13	h^2
1	-24	58	07	42	27	-23	-10	-24	-25	23	09	04	10	90
2	-65	39	32	-07	11	13	23	08	10	03	-17	-09	09	83
3	87	17	15	02	05	-05	-14	-04	02	03	-16	-03	-07	86
4	47	36	32	-51	13	-19	-04	-18	-13	-14	-06	10	06	85
5	-30	-53	37	-15	23	-05	17	-04	-21	-19	13	05	-09	72
6	68	-24	-13	04	38	28	-12	11	15	-09	-16	09	15	87
7	12	66	13	37	08	-04	16	-07	09	11	03	02	06	66
8	25	-25	17	61	-12	-24	11	05	22	-22	-11	05	-03	72
9	-16	-74	13	17	30	14	16	-10	08	-18	09	-10	-23	87
10	35	-49	-38	19	30	-16	14	-13	08	16	-18	-14	09	79
11	48	-40	43	-05	08	-03	05	-10	05	14	-02	11	07	63
12	-35	44	-16	52	16	07	07	-15	05	-13	07	10	09	71
13	-67	-03	12	-09	41	22	-17	-12	-13	22	07	-09	09	82
14	34	-23	23	29	-13	24	05	-12	09	19	19	08	11	49
15	-71	14	06	39	16	09	09	-08	10	08	06	-05	13	77
16	-66	-49	17	08	18	18	07	17	-07	03	-10	12	03	84
17	-53	27	48	-19	-06	-15	11	09	19	12	-12	-03	-12	73
18	-53	-49	45	12	-06	-07	02	-05	07	10	02	-13	02	78
19	11	-08	34	-63	22	05	04	06	05	11	-16	-11	-18	67
20	-44	13	52	-16	-10	-19	-10	21	08	21	17	04	12	70
21	-20	40	38	-09	-05	29	08	19	-31	25	07	08	-08	66
22	61	-22	50	-03	10	-07	04	13	08	16	04	16	-18	79
23	08	-46	59	-20	-12	02	03	04	08	-14	-04	07	20	70
24	73	-25	20	-24	14	-24	-21	21	-14	-11	-07	-10	03	91
25	45	28	14	-14	12	32	-19	-22	16	-14	-04	05	-07	57
26	-27	19	-18	-41	40	07	-16	-06	19	06	-16	19	-14	63
27	-60	15	33	35	-20	-16	09	-15	-19	-11	-30	15	08	88
28	43	64	09	-19	04	20	13	04	-15	-03	10	-17	11	77
29	24	20	30	33	-12	30	-23	-08	04	-05	06	05	-17	50
30	38	-22	37	36	-18	07	14	06	-17	-11	07	11	-07	58
31	09	41	34	12	06	04	-19	30	-10	-03	-09	15	-07	48
32	39	29	-14	10	30	-05	21	29	-15	09	10	07	-13	55
33	46	44	16	-08	25	-19	23	-20	08	-27	17	07	05	74
34	64	45	-24	09	16	02	27	19	-02	08	07	06	10	84
35	31	-45	-27	48	23	-14	-21	15	08	-07	24	-11	07	82
36	-12	18	18	13	14	-18	-32	28	25	04	24	16	-08	49

Whether a factor analysis succeeds or fails in contributing to psychological knowledge depends primarily, in the writer's opinion (3), on the principles and the skills employed in the rotation process. Guided by the principle of simple structure, we rotated the planes individually, avoiding any extreme obliqueness of axes; especially in the early stages and even in the final stages of it we permitted more factors than there were dimensions to our factor space. With a population smaller than that of our previous study, we accepted, at first, any loading through $\pm.10$ as lying in the hyperplane, though in the final, finer rotation we sought the maximizing also of loadings through $\pm.05$ in addition to maximizing those within $\pm.10$.

Previous experience in rating studies, using representatively scattered variables in the personality sphere, indicates that one may expect the number of variables falling in the hyperplane (for 12 factors and 36 variables) to vary from 15 to about 25. When less than 15 are obtained it is almost certain that the true hyperplane has not been found. On the other hand, protracted rotation causes such a factor to approach, but never to exceed, a limit of about 20 to 25 variables in the hyperplane. In other words, in personality material of this comprehensive kind, the criteria of at least one zero loading for each factor, which Thurstone adopted in the study of abilities, becomes insufficiently exacting.

In view of the importance for the whole field of personality study of the fact — discussed later — that the factors here obtained definitely confirm most of those of the previous study (1), it is desirable to emphasize that the factorization in this research was entirely independent of and unguided by that of the previous study. *The variables were represented in all tabular matter only by numbers and in graphs by unnumbered points, so that no hopes or prejudices of those engaged in the rotation could influence the direction of search for simple structure.*

The search for structure was rather a long one, taking the equivalent of one person's full time for a year. The difficulty of finding structure increases disproportionately with the number of factors and one does not expect an easy job with twelve; but for some reason* it proved rather more elusive here than in the previous study, in that tentative successes were often followed by dismays, renewed and re-renewed.

It is perhaps noteworthy that the factors which steadied first — *A*, *B*, *G*, *H*, *I*, and *K*—are those which most accurately reproduce fac-

* Possible reasons are the larger probable error of the r's due to the smaller population and the somewhat more uneven reliability of the ratings.

TABLE 2

Factors Rotated for Simple Structure
(In descending order of contribution to variance of all variables)

Variables	Factors 1	2	3	4	5	6	7	8	9	10	11
1	–03	0	07	–12	49	05	04	21	–47	0	–07
2	25	44	–02	–21	23	09	02	–02	–02	07	–12
3	–05	–20	–06	13	–07	01	–12	29	–16	–19	07
4	58	–05	03	10	–10	–09	–12	06	–24	–08	02
5	33	–01	04	–47	–05	–13	0	–05	04	15	–05
6	–09	–39	–31	–07	–04	03	–18	30	03	03	–01
7	0	29	–14	04	33	03	05	10	–09	0	0
8	–22	22	12	–08	01	27	–04	16	24	0	0
9	20	02	–08	–59	02	–04	02	06	18	0	–02
10	08	–32	–15	–26	23	39	–20	01	–24	22	01
11	–05	–17	08	–08	01	–01	02	02	–14	20	22
12	–04	25	–09	–08	34	05	–03	11	08	–05	–17
13	06	–13	01	–39	26	–12	12	10	–35	05	–08
14	–26	–09	–01	06	0	–23	10	–06	05	20	14
15	–09	28	02	–28	34	07	12	04	–05	11	–12
16	–03	–03	06	–33	16	10	–02	–03	10	20	0
17	31	55	24	–07	12	14	22	–03	0	–10	12
18	–07	16	33	–34	02	04	18	01	–08	15	02
19	39	01	04	–14	–04	–01	05	–02	–17	–09	17
20	–09	28	22	03	0	–09	45	01	–15	14	17
21	02	02	–04	06	10	–32	05	–07	–03	05	04
22	05	–04	–06	–03	03	–06	18	09	–03	01	37
23	18	03	19	–09	–29	–09	02	02	08	23	02
24	05	–37	–11	–08	–35	–04	–06	36	–19	–02	–05
25	15	–03	–06	11	–09	–22	–05	20	03	–32	04
26	35	04	0	01	31	20	08	–03	–11	–25	26
27	06	24	48	–03	23	27	26	02	01	04	–15
28	–16	0	–40	07	–14	–35	–10	07	–07	01	–24
29	–25	04	08	10	–14	–28	06	28	14	–32	03
30	–16	–05	0	–03	–14	–19	–11	08	21	09	–03
31	–05	04	04	09	02	–07	05	35	03	–19	02
32	02	–06	–51	–06	20	–03	01	05	0	05	05
33	51	25	–25	–07	06	–10	0	05	–02	–01	–05
34	01	–06	–54	14	13	–02	–08	–03	02	16	0
35	–45	–27	24	–24	–10	–01	17	36	–01	05	–09
36	–09	22	01	06	06	–04	53	29	01	–19	26

tors of the previous research. Factors *E* and *F*, which also showed a consistent character early (but not then labelled, of course), nevertheless proved quite difficult to settle to a really good hyperplane. The remaining factors, which took about a hundred more rotations, passing through numerous mutations of character, are those which actually match least the factors of the previous study.

After some 318 rotations, the simple structure of Table 2 was reached. Two of the original factors had meanwhile become residuals, leaving eleven with sufficient projection and consistency of projections and angles. Table 3 shows the direction cosines by which these are obtainable from the centroid analysis (Table 1) and the resulting angles among the factors themselves.

TABLE 3

Direction Cosines of Factors to Unrotated Factors and to One Another

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	–07	16	20	–57	32	–22	39	–30	04	–37	0	28	0
2	–27	30	24	05	–16	–18	41	03	64	–24	13	–04	–26
3	–23	–13	23	0	–47	–32	–40	–47	04	09	–32	22	–11
4	16	24	–23	–16	–66	01	–22	04	13	18	–04	54	07
5	–17	16	–06	25	45	–12	28	–25	10	53	–24	39	–14
6	–10	–07	–20	12	11	–48	14	0	28	15	–74	15	01
7	–12	03	16	–05	03	–18	–23	23	50	28	68	0	–15
8	12	09	25	33	38	–03	–61	22	–02	–30	–17	–36	04
9	–03	–07	–10	09	–35	34	28	31	24	–56	12	31	–27
10	–01	–22	02	–01	0	–05	51	12	–15	28	15	09	73
11	06	–08	05	–14	–03	–14	–02	–02	40	53	07	60	–39

	1	2	3	4	5	6	7	8	9	10	11
1											
2	33										
3	–02	02									
4	–15	02	40								
5	15	08	07	–07							
6	08	13	33	06	54						
7	–12	40	–07	08	–03	–29					
8	–30	–19	–07	–42	–26	–04	24				
9	05	45	–19	25	–47	–19	–01	–23			
10	06	–17	–30	–40	11	02	–09	–41	–17		
11	08	20	30	46	53	26	48	–44	04	–12	

5. *Description of the Factors*

Let us now examine the factors of Table 2 one by one, taking, as in the previous study (1), the highest five or six variables—those above a loading of about .25 — as a basis for recognizing each.

FACTOR E
(1 in this series)
Dominance vs Submissiveness
(Above .32) *Mean contribution to variance .048*

Variable	*Loading**	*Name and Direction of Variable*		
4	.58	Assertive, Self-assured	vs	Submissive
33	—.51	Secretive, Reserved	vs	Frank, Expressive
35	—.45	Independent-minded	vs	Dependent, Immature
19	.39	Hard, Stern	vs	Kindly, Soft-hearted
26	.35	Placid	vs	Worrying, Anxious
5	.33	Depressed, Solemn	vs	Cheerful

Variables in Hyperplane: Below ± .10 (Inclusive) *20*
Below ± .05 (Inclusive) *12*

* The sign of these loadings is that of the variable in *the polar direction as given in the original rating list.* For convenience of the reader the verbal labels have been reversed here where the variable has a negative loading, i.e., the sign applies to the variable number on the left, not to the verbal label to the right. Thus 33, Frank, was negative and has accordingly been reversed to read Secretive here.

The first factor in our series (in magnitude of variance) is clearly the well-known dominance pattern, labelled *E* in the original series. Throughout we shall adhere strictly to the letter label and descriptive label fixed in the original study wherever the factor in the present study is obviously the same as that defined in the original series. As pointed out in the survey (3) of earlier factorizations, the true dominance factor departs somewhat from popular, clinical type conceptions. It does so by having loadings, on the side of dominance, for the following: gravity or depression, embittered, hard qualities and restriction of self-expression. The phrase "strong, silent man" hits the exact factor pattern better than do current psychological discussions of "dominance behavior," which apparently confuse surgency with dominance.

FACTOR G
(2 in this series)
Positive Character Integration vs Immature, Dependent Character
(Above .28) *Mean contribution to variance .045*

17	.55	Persevering, Determined	vs	Quitting, Fickle
2	.44	Emotionally Stable	vs	Changeable
6	—.39	Responsible	vs	Frivolous
24	—.37	Emotionally Mature	vs	Demanding, Impatient
10	—.32	Intellectual, Cultured	vs	Boorish
7	.29	Attentive to People	vs	Cool, Aloof

Variables in Hyperplane: Below ±.10 (Inclusive) *18*
Below ±.05 (Inclusive) *15*

The first glance at this factor suggested it might be either *C* or *G*, for these are distinctly alike in pattern; but the emphasis on perseverance and intellectual training eventually inclines one to label it *G*. The finding raises the possibility that *C* and *G* should be considered a single factor. (In the original *C* factor, 1, the first five variables in order are: Persevering-Quitting, Stable-Changeable, Calm-Restless, Emotionally Mature-Demanding. Infantile, Austere-Frivolous. In *G* they are: Mature-Immature, Conscientious-Slipshod, Persevering-Quitting, Stable-Changeable and Thoughtful-Unreflective.)

FACTOR H
(3 in this series)
Charitable, Adventurous Cyclothymia vs Withdrawn, Heboid Schizothymia
(Above .30) *Mean contribution to variance .043*

34	—.54	Gregarious, Sociable	vs	Self-contained
32	—.51	Marked Interest in Opposite Sex	vs	Slight Interest in Opposite Sex
27	.48	Somewhat Unscrupulous	vs	Conscientious
28	—.40	Composed	vs	Shy, Bashful
18	.33	Adventurous, Bold	vs	Cautious, Retiring, Timid
6	—.31	Frivolous	vs	Responsible

Variables in Hyperplane: Below ±.10 (Inclusive) *21*
Below ±.05 (Inclusive) *12*

This is evidently the second of the two main schizoid factors of the original study, which we hypothesized to be the constitutional factor in schizophrenic tendency. It differs slightly from the pattern of the former study by bringing in conscientious, "good" character qualities in the schizothyme pole and sex interest, as well as by having lessened emphasis on hostility. This last modification may result from our greater success in separating the two schizoid factors more clearly (by specifically adding variables calculated to belong to one and not the other); for our tentative conclusion had been that the hostile schizoid characteristics belonged to Factor *A* and the withdrawing characteristics to Factor *H*.

The new incidence of "conscientious" characteristics is also in accordance with this constitutional interpretation, for studies of the pre-psychotic personality of the slow, insidious schizoid onset definitely stress the good, obedient, conscientious-in-school characteristics (3). Variable 32, Interest to Opposite Sex, could not have entered into our original pattern of *H* because it was not employed as a variable in that study. It was introduced here because it admitted of good ra-

ting (frequency of dating and sex conversation being readily observable in a residential fraternity). This variable also helps confirm the general interpretation of factor *H*, for deficient heterosexual interest has long been noted clinically to be part of the more constitutional schizoid pattern.

FACTOR F
(4 in this series)
Surgency vs Desurgency
(Above .30) *Mean contribution to variance .038*

9	—.59	Energetic, Alert	vs	Languid, Slow
5	—.47	Cheerful	vs	Depressed
13	—.39	Emotional	vs	Calm, Phlegmatic
18	—.34	Adventurous, Bold	vs	Cautious, Retiring, Timid
16	—.33	Talkative	vs	Silent, Introspective

Variables in Hyperplane: Below ±.10 (Inclusive) *22*
Below ±.05 (Inclusive) *6*

Considerable difficulty was encountered in getting a really good hyperplane for this factor, the numbers in the ±.05 range always turning out smaller than expected. Examination of the meaning of the many experimental rotations shows, however, that they all yield clearly the surgent-desurgent pattern containing, from the highest down, Cheerful—Depressed, Energetic—Languid, Talkative—Silent, Introspective. The best rotation we could get, above, (not necessarily the best obtainable if months could be spent on one factor), raises "energetic" above the usually highest loaded item "cheerful." It also stresses the "emotional" variable, though, as previous studies show, the emotionality of surgency is only of the cheerful, tender, sociable, and assertive forms, the emotionality of fear, anxiety, sorrow, and depression showing itself more in the desurgent pattern.

FACTOR A
(5 in this series)
Cyclothymia vs Schizothymia
(Above .26) *Mean contribution to variance .037*

1	.49	Ready to Cooperate	vs	Obstructive
24	—.35	Emotionally Mature	vs	Demanding, Impatient
12	.34	Good-natured, Easy-going	vs	Spiteful, Critical
15	.34	Mild, Self-effacing	vs	Self-willed, Egotistical
7	.33	Attentive to People	vs	Cool, Aloof
26	.31	Placid	vs	Worrying, Anxious
23	—.29	Adaptable	vs	Rigid

Variables in Hyperplane: Below ±.10 (Inclusive) *18*
Below ±.05 (Inclusive) *11*

In spite of some slight uncertainty over the exact hyperplane (the second poorest to *F*) this factor turned out to match the factor *A* of the previous research remarkably closely, six of the seven highest variables among the 36 being the same and occurring in the same order of magnitude. The additional variable here is 24, Demanding, Impatient, occurring at the schizoid pole. The three variables thrown into the factorization as likely, in view of the essence of the previous factor, most clearly to represent this source trait — namely, readiness to enter into cooperation, easygoingness, and freedom from rigidity of habit — have all emerged in the outstanding loadings.

FACTOR K
(6 in this series)
Trained, Socialized, Cultured Mind vs Boorishness
(Above .25) *Mean contribution to variance .025*

10	.39	Intellectual, Cultured	vs	Boorish
28	—.35	Composed	vs	Shy, Bashful
21	—.32	Polished	vs	Clumsy, Awkward
29	—.28	Sensitively Imaginative	vs	Practical, Logical
8	.27	Unshakable Poise	vs	Easily Upset
27	.27	Conscientious	vs	Somewhat Unscrupulous

Variables in Hyperplane: Below ±.10 (Inclusive) *23*
±.05 (Inclusive) *14*

K, as one of the slighter factors in the previous research, was in some doubt. Its emphatic hyperplane here, and clear matching of the previous pattern, bring it on a secure basis. Politeness and poise are a little more evident in this pattern, but the main basis of cultured interests, sensitivity and conscientiousness, remains.

FACTOR B
(7 in this series)
Intelligence (in Total Personality) vs Mental Defect
(Above .18) *Mean contribution to variance .024*

36	.53	Intelligent	vs	Stupid
20	.45	Insistently Ordered	vs	Relaxed, Indolent
27	.26	Conscientious	vs	Somewhat Unscrupulous
17	.22	Persevering, Determined	vs	Quitting, Fickle
10	—.20	Intellectual, Cultured	vs	Boorish

Variables in Hyperlane: Below ±.10 (Inclusive) *22*
Below ±.05 (Inclusive) *16*

This is clearly factor *B*. It is interesting to notice that by using

an actual intelligence test, instead of a pooled rating of intelligence as in the previous research, one gets a somewhat more definite hyperplane, though the pattern of "characterial" traits around intelligence is just about the same.

FACTOR I
(8 in this series)
Sensitive, Imaginative Emotionality vs Rigid, Mature Poise
(Above .28) *Mean contribution to variance .024*

24	.36	Demanding, Impatient	vs	Emotionally Mature
35	.36	Dependent, Immature	vs	Independent-minded
31	.35	Aesthetically Fastidious	vs	Lacking Artistic Feeling
6	.30	Frivolous	vs	Responsible
3	.29	Attention-getting	vs	Self-sufficient
36	.29	Intelligent	vs	
29	.28	Sensitively Imaginative	vs	Practical, Logical

Variables in Hyperplane: Below ±.10 (Inclusive) *26*
Below ±.05 (Inclusive) *15*

This has slight resemblances to Factor *D* of the original research, but the present writer and six psychologists who matched old and new factors independently, agree that the over-all resemblance is undoubtedly to Factor *I*. It is at this point, however, as one descends to the factors of least variance, that the first shadows of uncertainty over identification enter. Factor *I*, or anything resembling it, had previously been found only in one research (1), and we were accordingly doubtful about its nature. Since the hyperplane and variance of the present study are clear-cut and satisfactory, we are inclined to regard the above pattern as the more correct picture where it departs from the previous one, namely in dropping out "neurotic" and "emotional" from the *I*+ pole and substituting "aesthetically fastidious." It is interesting that in discussing the *I*+ pole previously it was illustrated by the case of an artist!

FACTOR J'
(9 in this series)
Thoughtful Neurasthenia vs Vigorous, Simple Character
(Above .20) *Mean contribution to variance .022*

1	—.47	Obstructive	vs	Ready to Cooperate
13	—.35	Independent-minded	vs	Dependent, Immature
4	—.24	Submissive	vs	Assertive, Self-assured
8	.24	Languid, Slow	vs	Energetic, Alert
10	—.24	Intellectual, Cultured	vs	Boorish
30	.21	Neurotically Fatigued	vs	

Variables in Hyperplane: Below ±.10 (Inclusive) *22*
Below ±.05 (Inclusive) *14*

This factor resembles only one of the original factors, *J*, to which, however, it is rather "off-center," in bringing "obstructive" and "independent" to the fore. The original *J* was a straight neurasthenia-like picture, showing lack of will qualities, fatigue, inhibition, meekness. It is quite likely, in view of the whole pattern, that the obstructiveness loaded here is the "obstructiveness" of a student who thinks for himself and goes his own way, though with much internal conflict. It may be the "Liking Thinking" factor of Guilford (7). In any case the relative emphasis on independent thinking in this student group and on neurasthenic fatigue and inhibition in our original, older group (30 years) might be due to a normal change of the pattern with age. It might well be called the "Hamlet Factor" in view of its loading the chief trait elements of that character. Tentatively, until research can clear the above point, we shall conclude that this is factor *J*, but to indicate that the matching is not as certain as the others we shall label it "*J*'" and call it Thoughtful Neurasthenia vs Vigorous, Simple Character, to retain true descriptiveness.

FACTOR M
(10 in this series)
Spiessburger Concernedness vs Bohemian Intellectualism
(Above .21) *Mean contribution to variance .020*

25	—.32	Conventional	vs	Unconventional, Eccentric
29	—.32	Practical, Logical	vs	Sensitively Imaginative
26	—.25	Worrying, Anxious	vs	Placid
23	.23	Rigid	vs	Adaptable
10	.22	Boorish	vs	Intellectual, Cultural

Variables in Hyperplane: Below ±.10 (Inclusive) *20*
Below ±.05 (Inclusive) *15*

This factor has no obvious counterpart in the original factors *A* to *L* (though certain rotations of it resemble *G*) and it will accordingly be listed as a newly discovered factor, *M*.

It may be a factor altering in variance a good deal from group to group, for it was absent in our last group, yet in others it is evidently great enough in variance to create a visible syndrome. The syndrome is that which has been called "Phillistinism" at one pole and perhaps "Bohemianism" at the other. The positive loading is very much like what the Germans used to call the "Spiessbürger" type — the correct, narrow citizen. Again it is that which has been typified by the antithesis of Martha and Mary. On the one hand the worthy but narrow soul, concerned to help in practical matters and have "no nonsense," doing the right thing; on the other, the individual neglectful of prac-

tical and community matters, unconcernedly living in intellectual interests and going his own way. One may speculate that the former will prove to have associations with number ability, the latter with verbal ability and higher perseveration. One psychologist, familiar with the rather marked difference between the rural midwestern and the eastern city representation in this student group has suggested that it is a rural-urban factor, but the writer believes it goes deeper than this and is concerned in any case to use at present a descriptive rather than interpretive label.

FACTOR L
(11 in this series)
Paranoid Schizothymia vs Sensitive, Trustful Accessibility
(Above .21) *Mean contribution to variance .016*

22	.37	Prone to Jealousy	vs	Free of Jealous Tendencies
26	.26	Placid	vs	Worrying, Anxious
36	.26	Intelligent	vs	Unintelligent
28	—.24	Shy,Bashful	vs	Composed
11	.22	Suspicious	vs	Trustful

Variables in Hyperplane: Below ±.10 (Inclusive) *23*
Below ±.05 (Inclusive) *19*

In spite of the small variance this factor has a very definite hyperplane, not permitting much change of meaning in the several experimental rotations tried. It contains both of the variables — Jealous and Suspicious — thrown into the pool to mark the paranoia factor *L*, which was indicated with only small and unreliable variance in the previous study. That it indicates the same paranoid schizothyme factor is shown also by the loading in shyness. On the other hand, the association with intelligence is unexpected and may be due to college entrance selection or other fortuitous influences in this sample. The title, but not the letter label, has been altered slightly in view of the clearer definition of the nature of the factor in the present study.

6. *Review*

Comparing this study of males averaging twenty years with the previous study of males averaging thirty years of age, we find that nine of the 11 factors now discovered are exactly identifiable with those of the earlier study, one is probably identifiable and one is quite new. It should be noted that this has occurred in spite of a different statistical process of handling the ratings in the two studies. The writer's former contention that the greater part of the variance in the personality sphere can be accounted for by about 12 factors, that these factors are stable over adult age ranges and that they have the

specific characteristics previously described, is therefore sufficiently confirmed. A defined foundation is henceforth provided upon which the search for objective tests of these factors can go forward, guided by the descriptions of the essential character of each factor.

On the other hand, there are certain puzzling and challenging discrepancies and shortcomings. Two of the factors found in the previous study — *D* and *C* — have vanished. It was remarked in the earlier study that *D* was most difficult of all to settle on a sound hyperplane, so its disappearance is least surprising. But *C* was well defined and has since been found also in a *P*-technique study (6). Its loss is a negation also of a part of the plan to bring out sharply the differences of similar, cooperative factors; for principally we aimed to distinguish the two cyclothyme factors, *A* and *H*, and the two "character" factors, *C* and *G*. The former succeeded well; the latter failed to the point of losing one of the factors concerned!

Before this is commented on further, the facts should be considered concerning the relative magnitude of the same factors in the two researches and the relative obliqueness of the same factors. The factors, which fell in alphabetical order before, now fall in the following descending order of contribution to variance — *E*, *C*, *H*, *F*, *A*, *K*, *B*, *I*, *J*, *L* — which correlates .45 (rank order) with the previous order. The greatest change is found in *B*, the next in the reversal of influence of the two cyclothyme factors, *A* and *H*, the next in the rise of *E* (Dominance). That *B* should fall is exactly what one would expect; for the normal range of intelligence variation would be at least halved in a college population. The rise of *E* may well be a true, systematic effect also, due to the greater role of dominance-submission relations in adolescent years. The reversal of the relative importance of the two cyclothyme-schizothyme factors is inexplicable unless we again adopt the age hypothesis and suppose, in accordance with the high adolescent incidence of schizophrenia, that the effects of the presumed constitutional factor *H* are exaggerated in this period. For the general range of adult life the magnitudes indicated by the original alphabetical labels are probably more nearly correct, and we do not contemplate departing from the original stable letter nomenclature there adopted.

Regarding the angles of obliqueness, showing the correlations among the factors, the agreement of the two studies is poor. This is no reflection on the identification of factors, because, as Thurstone has shown (9), these angles would be expected to alter with differences in sampling, especially in response to selection with respect to any variable. Further, the vagueness of some hyperplanes precludes determination of certain angles to within 10 or 20 degrees.

Poor agreement is concentrated in the angles of only a few factors: *A*, *F*, and *K*. Now *A* and *F* have the poorest hyperplanes (as seen if we add for each factor the number of variables within ±.10 to those within ±.05, as listed above). If we omit *F*, which was poorly fixed as to angle also in the previous study, the 28 angles among eight factors correlate +.02 with those of the previous study. The disagreement of the clearly hyperplaned *K* is strange, unless the variables in *K* are such as to come in for marked selection in a college population, a likely hypothesis. It is noticeable, however, that the correlations among the larger factors commented on in the summary of the previous study remain (save for *F*) largely the same. Thus *B* (Intelligence) and *C* (*G*) (Emotional Stability vs Immaturity) are appreciably positively correlated (formerly .32; here .40); *A* (Cyclothymia-Schizothymia) and *B* remain uncorrelated (*f* — .06; *h* — .03); the two cyclothyme factors, *A* and *H*, are faintly positive (*f* .11; *h* .07); *C* (*G*) and *E* (Dominance-Submissiveness) are positively correlated (*f* .39; *h* .33) and *B* and *I* (Sensitive Emotionality vs Tough poise) are positively correlated (*f* .36; *h* .24).

Returning to the one real hiatus in the confirmatory and clarifying function of this study—namely, the disappearance of factors *D* and *G*—two lines of research now suggest themselves.

(1) Our knowledge of the variance of variance of factors in different populations and conditions is still meagre, but the present findings, notably with the *B* factor, suggest that demotions of factors through selection can be sharp and that with relatively small populations (100 to 200) a factor for which the population is even moderately selected might easily fall to the position of a mere residual. Possibly the two residual factors we set aside during rotation are *D* and *C*. This could be answered by research with a large population. One can imagine college selection narrowing the range of *C* (Emotional Stability) considerably; but not of *D*.

(2) The special difficulties which these two factors present to discovery by rotation may have caused them to be missed despite being inherent in the data. The difficulties are those of poor hyperplane (It was pointed out in the previous study that rotation for *D* proceeded long after other factors were stabilized, because seemingly it leaves little in personality uninfluenced and creates no good hyperplane) and of extreme obliqueness (The correlation of *G* with *C* in the last study was .43, the largest of the 66 angles among factors). The existence of special difficulty is shown independently by the fact that in two alternative rotations of the previous study, *G* disappeared in one and both *G* and *D* in the other (2), the variance of *G* passing largely into *B*. These factors, therefore, have a special elusiveness,

and it would be worth while to enter the factor space again with trial vectors in an attempt to find them, guided this time by the obliquenesses found in the previous study. However, in view of the fact that the original work propounding the twelve factors has been under discussion for three years without crucial confirmation, it was thought desirable to publish these main findings before spending additional time searching for the two errant factors.

The writer wishes to express his appreciation of research assistants Dorothy Dibble, responsible for supervising 135 raters and compounding their results, and Ann Waldron, responsible for pursuing blind rotations to a definitive position.

REFERENCES

1. Cattell, R. B. The description of personality: principles and findings in a factor analysis. *Amer. J. Psychol.*, 1945, 58, 69-90.
2. Cattell, R. B. Simple structure in relation to some alternative factorizations of the personality sphere. *J. gen. Psychol.*, 1946, 35, 225-238.
3. Cattell, R. B. The description and measurement of personality. New York: World Book Co., 1946.
4 Cattell, R. B. A guide to mental testing. London: Univ. of London Press. Revised Edit., 1947 .
5 Cattell, R. B. Comparison of the primary personality factors in men and women. (In press).
6. Cattell, R. B.; Cattell, Karen A. S.; & Rhymer, Rue M. *P*-technique demonstrated in determining psycho-physiological source traits in a normal individual. *To be published in Psychometrika.*
7. Guilford, J. P. & Guilford, R. B. Personality factors, *D*, *R*, *T* and *A*. *J. abnorm. soc. Psychol.*, 1939, 34, 1.
8. Symonds, P. M. Diagnosing personality and conduct. New York: Appleton Century, 1931.
9. Thurstone, L. L. Primary mental abilities. Chicago: Univ. Chicago Press, 1938.
10. Thurstone, L. L. The effects of selection in factor analysis. *Psychometrika*, 1945, 10, 165-198.

PSYCHOMETRIKA—VOL. 12, NO. 3
SEPTEMBER, 1947

THE CONSISTENCY OF THE ITEMS OF AN ACTIVITY PREFERENCE BLANK

H. M. FOWLER
UNIVERSITY OF TORONTO

Results of an experiment to obtain data on the consistency of the items of two forms of an Activity Preference Blank are presented. Both Form I and Form II, which was a revised edition of Form I, were administered twice, so consistency data are available for both forms. A sub-item is said to be consistent if a high proportion of men marked it the same way, *M* for preferred Most and *L* for preferred Least, on both administrations of the test. The data of the experiment were investigated to see what happens to the consistency of sub-items when the items are changed in context, when the number of sub-items in an item is reduced, and when the time-interval between the administration and the re-administration of the test is increased. The author also gives data on the consistency of the responses made to particular combinations of sub-items and data on item consistency when all sub-item combinations are taken into consideration.

1. *Introduction*

One result of the impact of World War II upon the field of testing and measurement has been to show the very great need for adequate personality tests. There can be little doubt that the present personality tests leave much to be desired and that the war years have done a great deal toward bringing about this needed improvement. Almost every defence research project has found it necessary to do something about getting at the quirks of personality which cannot be reached by the all-important performance-instrument tests.

The present paper is based on an experiment done for the Applied Psychology Panel, NDRC, Office of Scientific Research and Development, in connection with research on the selection and training of navy radar and range-finder operators.*

2. *The Test Used*

The test used in the experiment was constructed for one phase of research concerned with differentiating personality types among high school juniors.† Slightly revised, this Activity Preference Blank

* Adams, J. K., and Fowler, H. M., *The Reliability of Two Forms of an Activity Preference Blank*, OSRD Report No. 3778, Applied Psychology Panel, NDRC.

† Fowler, H. M., *Preference-Ability Types of High School Junior Boys*. Unpublished doctorate thesis, 1945, Harvard Graduate School of Education, Cambridge, Mass.

was administered to beginning classes (about 800 men)* at the Naval Training School, Fire Controlmen (R), Fort Lauderdale, Florida. Form I was re-administered to two classes (approximately 200 men) to obtain data on consistency. This form is made up of 93 items each containing six sub-items. For each item the subject indicates the activity (sub-item) which he prefers most and the activity (sub-item) which he prefers least. Those sub-items which were found to be consistent were used in constructing Form II, which is made up of 84 items each containing four sub-items. Form II was re-administered to one class (approximately 100 men) to obtain data on its consistency.

The following are some examples of items used in the test:

2. a. Help to dig a ditch.
   b. Fix a "flat" on the car at home.
   c. Make an outline of the principal points learned in the lessons of the week.
   d. Cut the lawn at your home.

22. a. Fix a radio.
   b. Take a lesson in first aid.
   c. Go to hear a famous politician speak.
   d. Read about the causes of various diseases.

40. a. Read "Popular Science" magazine.
   b. Look at prints for sale in a nearby art store.
   c. Attend a murder trial.
   d. Watch a skilled craftsman at work.

3. *The Preference Item*

The chief characteristics of the preference item are:

1. It calls for a comparison of one sub-item against another.
2. The responder ranks sub-items in a certain order. If there are more than two sub-items in an item, usually the ranking consists merely of selecting the sub-item liked least and that liked most.
3. The preference item has the advantage of yielding considerable information in a compact form. For example, when a "least" and a "most" response are made to a four-choice item, the results of five independent comparisons appear in the one marking.
4. The responder does not refer to an absolute scale of liking as in the case of the like-dislike item. This allows comparisons of sub-items, all of which may be liked or disliked by the responder. (See first example above.)

* The men involved in the experiment came directly from basic training school. They varied widely in ability, background and, presumably, in interests. Most of them were young men in their early twenties or late teens.

Lee and Thorpe* argue as follows: "When items are presented singly, it is often difficult for the individual to decide whether he likes, dislikes, or is neutral regarding a given choice; these terms are somewhat vague and provide no point of reference for comparison. On the other hand, when two groups of activities are presented in contrast, the situation becomes more concrete and meaningful and contributes to a more accurate choice. Research appears to confirm the superior stability of choices made in this manner. The above method also requires the individual to choose between two groups of activities even though he is interested in neither or both."

"Research appears to confirm the superior stability of choices made in this manner." This statement refers to the work done by Kuder,† who has indicated that the preference item is very stable and that it yields as reliable results as the like-dislike item.

4. *Method of Determining Consistency*

For purposes of this article, a "consistent" item is defined in general as one which a high proportion of men mark the same way for two different administrations of the test. The frequency with which each sub-item was marked *M* (most preferred) and *L* (least preferred) in each of the 84 items of the blank was recorded for both the administration and the re-administration of the test. From these data, two values were obtained for both *M* and *L* responses to each sub-item.

TABLE 1

Sub-Item Consistency Data

Item	Sub-Item	Activity	Consistency Data	
			M	L
	a	Coach your fellows for a sport in which you are proficient.	57 - 9	1 - 1
	b	Build an attractive cabinet out of hardwood.	17 - 6	5 - 4
	c	Try out cooking recipes.	1 - 0	40 - 15
	d	Work in an office during the summer holidays.	5 - 3	21 - 13

* Lee, Edwin A. and Thorpe, Louis P. *Occupational Interest Inventory. Advanced Series, Manual of Directions.* Los Angeles: California Test Bureau, 1943, p. 6.

† Kuder, G. Frederic. The stability of preference items. *J. soc. Psychol.*, 1939, 19, 41-50.

For example, consider the data shown in Table 1 for the first item of the test.

For sub-item "a" of item 1, 57 men marked the *M* response on *both* administrations of the test; nine men marked *M* for sub-item "a" on the first administration but made some other response on the second administration. In other words, this sub-item appears to be consistent for the *M* response because a high proportion of men mark it the same way for both test and re-test. Two questions immediately arise: (1) Could we expect this result to happen by chance? (2) Is the proportion of men who marked the sub-item the same way on both administrations *significantly* greater than the proportion who failed to do so?

The investigation of these questions led to the setting up of two criteria for consistency. For a sub-item to be consistent: (1) the frequency of the *M* (or *L*) response for both test and re-test must be at least 15; (2) the ratio of the number of men marking the same way and the number failing to do so must be 2/1 or greater. These criteria were arbitrarily selected so that the probability of obtaining the result by chance would be .01 or less.

With an item made up of four sub-items, the probability that *n* out of 100 individuals will respond *M* (or *L*) to a given sub-item on *both* test and re-test *by chance* is given by the equation:

$$p = \left(\frac{1}{16}\right)^n \left(\frac{15}{16}\right)^{100-n} \frac{100!}{n!\,(100-n)!}.$$

With this equation it is easy to calculate that the probability that one out of 100 men will mark by chance the same sub-item on both test and re-test is .989507. In other words, it is very likely that at least one individual out of a group of one hundred will mark the same sub-item merely by chance. But the total probability that fourteen, or fewer, out of one hundred individuals will do so is only .01264. Thus when we say that the frequency of *M* (or *L*) response for both test and re-test must be at least 15 we are demanding that our results must be significant at the one per cent level. The ratio, 2/1, was chosen as the criterion of consistency for the proportion of *similar* response on both administrations to *different* response on the second administration because this ratio is also significant at the one per cent level for the minimum *N* which satisfies criterion one. This may be easily determined using the usual formula for the significance of the difference between two proportions for an *N* of 22 — the smallest *N* satisfying both criteria, 15 giving a similar response on both administrations and seven giving a different response on the second administration.

5. *Consistency of Sub-Items Under Changed Conditions*

Kuder* has stated that "the greater the number of parts to be ranked in a group the less is the time and attention given to each preference made and, it would seem, the less reliable would be each preference involved." In this experiment data are available for:

(a) different conditions of context
(b) number of sub-items in an item
(c) time interval between test and re-test

a. *Context.* This refers to the way sub-items are grouped to form items. Context is particularly important in a preference test because the responses for a sub-item depend not only upon the responder's liking for that activity but also upon his liking for other activities in the same item. Eight items of Form II were constructed by grouping miscellaneous sub-items of Form I; in other words, their sub-items are in different context. For example, item 54 of Part I of Form II is as follows:

a. Go to see George Arliss play in the "Merchant of Venice."
b. Go deep-sea fishing with a group.
c. Travel on a lake by means of an outboard motor.
d. Take part in an after-dinner discussion.

Sub-items "a" and "d" were obtained from item 20 of Form I, sub-item "b" from item 22 and sub-item "c" from item 34 of Form I. A small number of items of Form I were, in general, inconsistent, but consistent sub-items were taken from them and re-grouped to form new items in Form II. The majority of items of Form II, however, have the same general context as they had in Form I.

b. *Number of Sub-Items in an Item.* Form II items have four sub-items whereas Form I items have six sub-items. In general the items of Form II are the items of Form I with two sub-items omitted from each item. In a way this constitutes a change of context but it is considered in this article to be a reduction in the number of sub-items. For example, in Form I, item 47 is as follows:

a. Go on a long hike.
b. Be a model in a fashion show.
c. Play in an important football game.
d. Read the society page of the local newspaper.
e. Listen to the news on the radio.
f. Go to a meeting of a Bible class.

This appeared as item 27 in Part I of Form II as follows:

a. Go on a long hike.

* Kuder, G. Frederic, *op. cit.*, p. 44.

b. Play in an important football game.
c. Listen to the news on the radio.
d. Go to a meeting of a Bible class.

Two of the sub-items of the original item were omitted because it was found that very few boys, for whom Form II was intended, selected them either as "most" or "least."

c. *Time Interval Between Test and Re-Test.* Form I was administered twice to one class of men at an interval of two weeks and to another class twice at an interval of eleven weeks. Form II was administered to one class of men and then re-administered after a period of three weeks. For purposes of this discussion the two-week and three-week intervals are considered to be similar time intervals.

TABLE 2
Consistency of Sub-Items at Different Time Intervals

Time Interval Between Test and Re-Test	*M* Response		*L* Response	
	No. of Consistent Sub-Items	Total	No. of Consistent Sub-Items	Total
Two weeks	129	558	38	558
Eleven Weeks	91	558	42	558
	Contingency Coefficient $r = .69$		Contingency Coefficient $r = .62$	

Tables 2 — 6 show what happens to the consistency of the sub-items under various conditions.

For both the *M* response and the *L* response there is a high relationship between the consistency of the sub-items at a two-week test-re-test interval and their consistency at an eleven-week interval (Table 2). In other words, if a sub-item is consistent at a two-week interval it tends to be consistent also at an eleven-week interval. For the two-week interval between test and re-test there are over three times as many consistent sub-items for the *M* response as for the *L* response (129 to 38). When the interval is increased to eleven weeks there are still over twice as many consistent *M* responses (91 to 42). The increased time interval reduces the number of consistent sub-items for the M response from 129 to 91; for the *L* response, however, there is no such reduction but actually a slight increase from 38 to 42.

For both *M* and *L* response a high relationship exists between the consistency of sub-items in a six-choice item and their consistency in

TABLE 3
Consistency of Sub-Items for Items of Different Length

Number of Sub-Items in Each Item	*M* Response		*L* Response	
	No. of Consistent Sub-Items	Total	No. of Consistent Sub-Items	Total
Four	124	304	54	188
Six	118	304	26	188
	Contingency Coefficient $r = .69$		Contingency Coefficient $r = .69$	

a four-choice item (Table 3). This means that increasing or decreasing the number of sub-items, to this extent at least, does not greatly affect the consistency of individual sub-items. Note: the only change of context was the omission of two sub-items with low response frequencies.

There is a higher proportion of consistent sub-items for the *M* response than for the *L* response for both the six-choice item and the four-choice item. However, the difference in the proportions is much greater for the six-choice item (see next observation).

The reduction of the number of sub-items from six to four more than doubles the number of consistent *L* responses (from 26 to 54) but only slightly increases the number of consistent *M* responses (from 118 to 124).

For both the *M* and the *L* responses there is a high relationship

TABLE 4
Consistency of Sub-Items at Different Time Intervals *and* for Items of Different Length

Time Interval and Number of Sub-Items in Each Item	*M* Response		*L* Response	
	No. of Consistent Sub-Items	Total	No. of Consistent Sub-Items	Total
Three Weeks and Four Sub-Items	124	304	54	188
Eleven Weeks and Six Sub-Items	81	304	31	188
	Contingency Coefficient $r = .65$		Contingency Coefficient $r = .70$	

between the consistencies of the sub-items even when both the number of sub-items per item is changed *and* there is a considerable difference in the time-intervals between test and re-test (Table 4). The relationship is slightly higher for the *L* response. For both six sub-items at eleven weeks and four sub-items at three weeks there is a higher proportion of consistent sub-items for the *M* response than for the *L* response. Reduction in the number of sub-items per item together with a decrease in the time interval between test and re-test increases the proportion of consistent sub-items for the *M* response from 27% (81/304) to 41% (124/304); for the *L* response the increase is from 16.5% (31/188) to 29% (54/188).

Table 4 is like Table 3 except for the addition of the time factor. As would be expected from our observations for Table 2, this addition brings about a reduction in the number of consistent sub-items for the *M* response (from 118 to 81) but a slight increase in the number of consistent sub-items for the *L* response (from 26 to 31).

TABLE 5
Consistency of Sub-Items When Placed in Items of Different Length *and* Different Context

Number of Sub-Items and Context	*M* Response		*L* Response	
	No. of Consistent Sub-Items	Total	No. of Consistent Sub-Items	Total
Four Sub-Items in One Context	12	32	6	32
Six Sub-Items in a Different Context	5	32	5	32
	Contingency Coefficient $r = .28$		Contingency Coefficient $r = .59$	

For the *M* response there is no significant relationship between the consistencies of the sub-items under different context (Table 5). This is due to the fact that many sub-items that were not consistent in the one context are consistent when put in a different context. (An increase from five to 12). For the *L* response the relationship is significant (at the 1% level of confidence). This is due to the fact that for this response sub-items tend to "stay put." Note that most of the sub-items that were inconsistent in the one context are still inconsistent when put in different context (the increase in the number of consistent sub-items is only from five to six).

It is clear that a change in context is of much more significance for sub-item consistency than a reduction in the number of sub-items per item or a change in the time interval between test and re-test. Although the data available in this experiment are not sufficient to warrant any final conclusions, it would appear that when the context is changed we can not be sure what will happen to the consistency of the sub-item. This is particularly true of the *M* response for which consistency seems to be more sensitive to changes in conditions.

For the *M* response there is no significant relationship between the consistencies of the sub-items under different context and changed conditions of number of sub-items and time interval between test and re-test. Comparison of the *M* response in Table 6 with the *M* response in Table 5 shows that the addition of the time factor as a variable makes very little difference. Context is relatively much more important, with respect to its influence upon consistency, than either the time factor or the number of sub-items in the item. For the *L* response the relationship between consistencies is high even when context, number of sub-items, and time interval are all variables. For some unknown reason (probably small-sample errors) the relationship is even higher in Table 6 than in Table 5.

TABLE 6
Consistency of Sub-Items Under Changes of Item Context, Item Length, and Time Interval

Compared Conditions	*M* Response		*L* Response	
		Total		Total
One Context for Four Sub-Items at Three Weeks	12	32	6	32
A Different Context for Six Sub-Items at Eleven Weeks	7	32	8	32
	Contingency Coefficient $r = .30$		Contingency Coefficient $r = .77$	

6. *Sub-Item Consistency*

A large number of the sub-items of both Form I and Form II satisfied the criteria of consistency discussed above.* Of 93 items,

* For Form I the frequency of the *M* (or *L*) response for both test and re-test needed to be only eight for significance at the 1% level. This was due to the fact that there were six sub-items in each item of Form I, whereas there were only four sub-items in each item of Form II.

each containing six sub-items, in Form I, 80 items with 100 sub-items gave consistent *M* responses but only 33 items with 34 sub-items gave consistent *L* responses. Of the 55 items in Part I of Form II there were 84 sub-items to which consistent *M* responses were made, but only 55 sub-items to which consistent *L* responses were given. Each item had at least one consistent *M* response but five of the items had no consistent *L* response. Of the 29 items in Part II, for which only *M* responses were asked, 26 of the items had 37 sub-items to which consistent responses were made.

The outstanding fact to be noted from these results is that the *L* response shows less consistency than the *M* response.

7. *Item Consistency*

Most of the interest and preference tests of the past have used sub-items for their scoring units. Scoring weights have been assigned to sub-items in accordance with the degree of relationship between these individual units and some criterion. Total scores for the tests have been simply the algebraic sums of these weights. A recent development in the interest testing field has been the suggestion that more useful scores might be obtained by devising some means of scoring by patterns of responses rather than by single responses. For a preference test the simplest pattern response is the combination within a single item of the *M* response and the *L* response. Item consistency refers to the consistency of such combined responses.

If an item has four sub-items, "a b c d," a response of *M* for sub-

TABLE 7
Consistency of Certain Combined Responses

Item	Sub-Item Combination	Consistency Data
1	ac	22 - 13
2	ca	(17 - 8)
3	ab	(38 - 15)
4	da	(39 - 9)
5	ad	(15 - 8)
	bd	(7 - 3)
6	ca	(5 - 1)
	bc	(25 - 15)
7	—	No consistent combination
8	ca	(6 - 1)
	bd	(6 - 2)
etc.		

item "a" together with a response of *L* for sub-item "c" may be thought of as a combined response, "ac." It is clear that there are 12 possible combinations for any item. The data of this experiment were investigated to determine the consistency of frequently-occurring combinations* for each item and the consistency of the items taking into consideration all *M-L* responses.

It was discovered that almost every item had at least one consistent combined response. Some examples of the data are given in Table 7.

Table 7 is to be interpreted as follows: the first number shown in the third column is the number of men who responded *M-L*, as shown in the sub-items combination column, on *both* administrations of the test; the second is the number of men who responded *M-L*, as shown in the sub-item combination column, on the *first* administration but who failed to make exactly the same response on the *second* administration. Those couplets which are enclosed in parentheses are ones which were consistent at the 1% level.

It should be remembered that the combinations shown in Table 7 account for only a part of the total response to each of the items. For example, in item 1, those who responded *M* for sub-item "a" and *L* for sub-item "c" made up less than half of the group being tested. What, then, is the consistency of the item when all combinations are taken into consideration?

Data for whole-item consistency were obtained from two groups, each of slightly more than one hundred men, who received both test and re-test of Form I. For the most consistent item, 105 men, or 48%, made exactly the same response for both *M* and *L* in both administrations of the test. Only 16% of the men made exactly the same response for the least consistent item. This shows that the items vary a great deal with respect to their consistency.

8. *Summary*

This paper presents data on the consistency of sub-items and items of an activity preference blank used in a war research project concerned with methods of selection of service personnel. It is noted that:

1. For the *M* response, the consistency of sub-items remains remarkably stable regardless of how they appear in items. Number of sub-items and time interval between test and re-test make very little

* For significance at the 1% level, the frequency of the combination response must be at least four. As would be expected, this is smaller than the frequency required for a single sub-item.

difference. The only thing that radically alters consistency of sub-items is a complete change of context.

2. For the *L* response, change of the time interval makes very little difference in consistency but reduction in the number of sub-items produces more consistent sub-items (see 3 below). The limited data of this experiment bearing on context suggest that this factor has little effect on the consistency of the *L* responses.

3. Kuder's conclusion that "the greater the number of parts to be ranked in a group the less is the time and attention given to each preference made and, it would seem, the less reliable would be each preference involved" finds support in the data of this experiment. However, a great difference is noted only for the *L* response.

4. The great majority of the items have at least one sub-item with a consistent *M* response (like Most).

5. A smaller number of sub-items have consistent *L* responses (like Least).

6. In general, *M* responses are more consistent than *L* responses.

7. Almost every item of the test has a certain *M*-*L* response which is consistent.

8. When consistency is based upon the responses to *both M* and *L*, in other words, when consistency is considered in terms of items rather than sub-items, great differences between items are found. The proportion of men repeating both responses ranges from 48% for the best item to 16% for the poorest item.

SEPTEMBER, 1947
PSYCHOMETRIKA—VOL. 12, NO. 3

BOOK REVIEWS

RAYMOND B. CATTELL. *Description and measurement of personality.* Yonkers-on-Hudson: World Book Company, 1946. Pp. ix + 602.

The Measurement and Adjustment Series of the World Book Company, under the editorship of Lewis M. Terman, gives us a new volume with the title above. It is a large volume, with a bibliography of 286 titles.

This is something to welcome, indeed, for nothing approaching this scope has appeared since Gordon Allport's *"Personality: A Psychological Interpretation."* That came out in 1937 and thus followed after the earlier impressionistic writers such as Jung, but before the rich surge of statistical inquiries which followed. It is high time that such a book as Cattell's appeared and that it should be that author to write it, because his work has been broad enough to do what is needed, viz: bring together the results from behavior ratings, personal questionnaires, objective tests, and clinical observations. It is also important that it should be done by one devoted to rigorous statistical methods, especially factor analysis.

As the preface warns us, however, the title given is not really the title for this book, but rather for a series of two volumes of which this is the first. A more modest and accurate title for this volume would be "The Classification of Temperament." The next will be on the development of temperament. There will be little help in this volume, and probably in the second, for one who wants to measure personality. For that, one must look to the new Buros handbook now in preparation, and one will obtain some help from Rapaport.

One shortcoming of the book is the fact that some sections of the book were obviously written long before others, so that there is not a fresh contemporaneity in every part. As an illustration, no papers by Guilford since 1939, by Humm since 1933, and by Hathaway none at all are cited. Fortunately, the volume to follow can make some amends; but probably not enough, since it is to deal specially with development.

In spite of this limitation, it is a very valuable achievement to collate the studies in four fields. Let us hope that the author will favor us periodically by gathering these studies, especially those on objective tests, since summaries of this type are so meagre and so greatly needed. So far, the results of objective testing are few and fit together poorly. Such work will take years to make a coverage comparable with the other approaches.

To the methodological psychometrist, the writer's discussion of factor-analysis techniques is likely to claim special interest. While somewhat eclectic, he makes much use of the Thurstone centroid method and of his own "principle of proportional profiles."

The book is to a great degree devoted to discussions of nomenclature. The word temperament is used more narrowly than by many students of personality. It is limited to behavior tendencies that are devoid of the dynamic, that is, that are the result of a drive rather than of a purpose, and so it is dependent upon physiological characteristics of the individual. It will be difficult to narrow

thus the use of the word temperament for those who wish to include attitude and interest as well as other behavior tendencies in the term.

Objection may well be made to the stress the author places on the terms "surface traits" and "source traits," meaning observed traits and those statistically derived by factor-analysis studies. As new factor-analysis results appear, "source traits" become adopted by test devisers as observed traits, so the distinction becomes less and less significant.

The strong preferences that the author shows for behavior ratings and objective testing as against the personal questionnaire are not supported by experience. It is reported that the U. S. Army Air Forces found as much validity for questionnaire scores for prediction of the pilot pass-fail criterion as for all the behavior ratings and objective tests (except personal-history data) taken together.* This finding should do much to alter the author's prejudice. He has given undue weight to the sources of error in the questionnaire and inadequate attention to very grave limitations in the other methods. Unquestionably, all three methods will be and should continue to be used, since none can be safely discarded.

Insufficient recognition is given to the validating procedures in the use of questionnaires, notably Hathaway's "K" score, which on further development will put questionnaire work on a better basis.

The author finds little merit in the projective methods and is indeed quite severe in his estimation of the leading test of this sort, the Rorschach ink-blot test. It is worth quoting his estimate (pp. 463 & 534). "Despite the attraction which the Rorschach test seems to have exercised—especially on psychiatrists—and the resulting great expenditures of time upon it, its validity and reliability seem to be little improved since its defective psychometric construction and poor level of prediction were reviewed in 1936. It remains a mixture of ill-defined intentions, analogous to a patent medicine, devoid of clear-cut theoretical bases which would permit refinement through research. No factor or cluster has been established among its separate measurements, which are defined in terms of highly subjective and *a priori* notions. One of the most recent studies (Sanford) shows its predictions to be 'slightly but reliably above chance.' But this average is achieved because 'intellectual power, special abilities, work habits were more reliably predicted than temperament (introversion-extraversion) or motivation.' In short, it succeeds in measuring abilities for which there are already more valid tests, rather than dynamic and temperamental traits in which research should now be interested. Thurstone's recent study, factor analyzing the Rorschach with other perceptual measures, found the factor to be specific to the test, not carrying over to other situations . . . . . The Rorschach . . . (deals) . . . with no well established variables and has lower validity than many less-complicated devices."

The author ambitiously presents a scheme based, in its first form, on his study of collected behavior ratings which he hopes may bring more agreement in the classification of temperament than now prevails. There would be more hope of getting fair agreement if it were based statistically on behavior ratings and questionnaire studies, but it is also based upon objective test studies and clinical impressions. With this diversity of sources, the conflicts and disparaties become so great that the result is largely the author's guess rather than an objective finding. For this reason, the present classification has no chance of

* Guilford, J. P. (Ed.) *Army Air Forces Aviation Psychology Program Research Reports*, No. 5. Washington, D. C.: Government Printing Office, 1947.

adoption as it stands. It is useful as a basis of discussion, however, and as something to be given statistical test.

Concerning this classification system the reviewer offers a few comments:

(1) The inclusiveness is excellent. One will see among the adjectives listed in the summaries nearly all the traits one is likely to find or need.

(2) The greatest fault is a too ready willingness to set up a pair of traits as opposite poles, where one pole is something more than a lesser degree. It is better to let assumed polarity await adequate correlation studies. The outstanding case is cyclothymia vs. schizothymia. Each of these poles is a well accepted trait, but there is not such a high negative correlation between them as to justify a polarity. One finds schizoids with what he calls cyclothmia, those with middling scores, and those with low scores. There is a negative correlation, but not the correlation close to —1.00 needed for polarity.

(3) In cyclothymia, as described by Cattell, the cyclic quality is one of its least important. In fact, it is quite swamped by other traits centering around cordiality. It would be better renamed.

(4) The trait "dominance-ascendance vs. submissiveness" as described is dual. On the one hand is included a mere activity, the opposite of which is not submissiveness. With this is the dominant-ascendant trait which is truly aggressive, aimed at another person. It would seem best to treat this aggressive-toward-others as the pole opposite to submissiveness and to release activity as a separate trait as is done by Guilford in his "general drive."

All in all, we have in this book a very valuable contribution, which will aid significantly in the next steps to be taken in this important field. It is a real service to step up the progress so much needed in the classification and measurement of temperament. We may well all look forward with eagerness to the volume to follow on the development of personality.

Hollywood, California ROSWELL H. JOHNSON

ALLEN L. EDWARDS. *Statistical analysis for students in psychology and education.* New York: Rinehart and Company, Inc., 1946. Pp. xviii + 360.

It is the author's intention to present an integrated introductory course in fundamental statistics employed in psychology, in education, and in the social sciences which emphasizes an understanding and appreciation of the tool value of statistical procedures and minimizes the objective of computation for its own sake. Designed to meet the modest mathematical backgrounds of most undergraduate students majoring in the social sciences, the text is definitely "applied" in its orientation. No assumptions are made concerning the reader's mathematical training. In order that one's understanding of subsequent topics may be assured, the author wisely has inserted an early chapter in which a survey of elementary principles of high school mathematics is presented. Derivations are introduced in those instances in which they add clarity to the development of a concept; for example, the derivation, in part, of the estimate of the standard error of a mean.

Clearly evident is the emphasis upon small sample theory in harmony with the current trend in published research. The author presents in a lucid style two fairly comprehensive chapters upon analysis of variance—a distinct contribution to the subject matter of text books in psychological and educational statistics. Numerous illustrative examples for both independent and matched groups serve to clarify the applications of analyses of variance.

Other topics in the text include the familiar measures of central tendency and variability, the product-moment correlation coefficient, and such well-known statistics of association as the correlation ratio and coefficients of biserial correlation, tetrachoric correlation, phi, contingency, and rank-difference correlation. An entire chapter is devoted to the chi-square test of significance with the same stress upon the testing of the tenability of hypotheses as that found in the chapter upon the *t*-test of significance and in the two chapters upon analysis of variance. Among the best written and most helpful of the chapters is the concluding one upon research and experimentation, in which problems of experimental design are treated. In another meritorious chapter, "Predictions and the Evaluation of Predictions," familiar topics of regression are considered. Particularly clear in its presentation is the chapter concerned with elementary probability and the binomial and normal distributions.

Although the author states in the preface his reasons for omitting an extensive treatment of partial and multiple correlation, the reviewer believes that the multiple regression equation and associated techniques are among the most valuable and most useful tools in psychological testing. The author's contribution of only one and one-third pages to the topic is unfortunate. Even if these techniques should be reserved for advanced courses, as the author suggests, a reasonably thorough consideration of them should be present if only for reference purposes. Much needed is an elementary, but adequate, discussion of reliability and validity of measures. Omission of consideration of the fallibility of measures is likely to leave the beginner with a distorted and unrealistic grasp of the actual precision of educational and psychological measurement. Some of the more recent trends in thinking upon the reliability and validity of tests might well have been included.

A helpful feature to both teacher and students is a set of examples at the conclusion of each chapter along with answers. Another aid is the comprehensive bibliography of one hundred and two references. Two tables not frequently found in current texts include one of values of the epsilon-square test at the five per cent and one per cent levels of significance and one of random numbers. The simplifying procedure of coding is stressed from its initial presentation as a major topic in the fourth chapter throughout the remainder of the text.

Technical errors are minor and few. Two minor slips which the reviewer observed were first a questionable procedure, or at least a misleading one, employed in ascertaining the fiducial limits for a proportion (see page 169) and second an inaccurate statement concerning σ_r (page 185). A technical point of considerable interest which is discussed at length (pages 179-180) is that of the number of degrees of freedom in the estimate of the standard error of a difference between means of samples in which the individuals of one sample have been paired with those of the other.

Although *Statistical Analysis for Students in Psychology and Education* is forward-looking in its content, the omission of certain "traditional" topics, the value of which is still apparent, is regrettable. Familiar texts in the field by Garrett, Guilford, Lindquist, and Walker appear to be better oriented to most introductory courses in statistical method than is Edwards' book. The reviewer does recommend highly, however, its use as a reference aid and as a source of exercises.

The University of Southern California WILLIAM B. MICHAEL

BERTHA B. FRIEDMAN. *Foundations of the measurement of values.* New York: Bureau of Publications, Teachers College, Columbia University, 1946. Pp. viii + 227.

The purpose of this book was to set forth the results of the author's search for instruments which will accomplish the task of measuring values to a useful extent. The importance of this problem is emphasized by the recent trend in education toward character training, personality development, and the development of desirable behavior tendencies in the school. The determination of an individual's values is a necessary preliminary and the basis upon which the educator must direct conduct.

Early chapters are devoted to a philosophical discussion of value theory, a discussion of the theories of motivation, and finally to a definition of value. The author defines value-behavior as "consummatory activity of any kind or as activity which is expected to bring about conditions favorable for consummatory activity."

Three possible sources for locating a person's values are indicated: (1) past and present overt behavior; (2) verbal statements; and (3) a combination of the two. None of these proves to be infallible as an index of future value-behavior and considerable disagreement may occur among them.

The concept of measurement is used rather broadly so as to include both location and quantification of values, since valid quantification depends upon proper location, the author believes. She also uses the term quantification in a broad sense, to include all numerical assignments that have a meaning of more or less than some other numerical assignment.

The quantification of values is discussed for the situations in which an individual has a certain amount of time, money, or energy to expend upon certain values, and for the situation in which he chooses between two alternatives. Value behavior itself may be recorded in units of time, money, or energy expended or as a ratio of the number of times one object is preferred to another. In deriving a value score, however, the further step must be taken of recording value-behavior over a period of time, and then deriving a quantity which is a function of the whole series of obtained magnitudes.

Most of the remaining chapters in the book are devoted to a description and critical discussion of various types of scales used in value measurement, including verbal scales, possible continuum scales, the scale of inclusiveness, and scales related to the balance of alternatives.

The greatest value of this book perhaps lies in its description of the various techniques which are available for measuring values. It is also valuable as a review of the literature in this field and for its account of the difficulties involved in this type of measurement.

It is probably least valuable as a discussion of the theory of measurement applied to this particular problem. The definition of value-behavior is phrased and expounded in such a way as to obviate the possibility of obtaining very accurate measurement in this field. She states that estimation by the measurer will always be necessary in value measurement due to the dynamic quality of the personality.

The book will be of most help to those persons who are actively engaged in the task of measuring values, attitudes, and various aspects of the personality. For those who are interested in the problem of developing psychological measurement on a rational basis, it will be of less interest.

The University of Southern California ANDREW L. COMREY